

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 12, 2002, 07:52:43 ; Search time 41.28 Seconds
(without alignments)
3366.117 Million cell updates/sec

Title: US-09-995-749a-2_copy_531_1781

Perfect score: 6641

Sequence: 1 MSLTARDVAIVPSLYNSA.....SDAEYPTSDVGRKMDQNK 1251

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : A_Geneseq_032802.*
1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*
22: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2690	40.5	2057	21	AA10667 L. mesenteroides a
2	2274	34.2	1592	14	AA32925 Glucosyltransferase
3	2175.5	32.8	1577	17	AA391047 Alpha-D-glucosyltr
4	326.5	4.9	349	21	AA36877 Glucosyltransferase
5	252.5	3.8	2314	22	AA369136 M. catarrhalis les
6	249	3.7	1349	22	AA34402 Staphylococcus aur
7	249	3.7	1349	22	AA37544 Staphylococcus aur
8	246	3.7	2893	19	AA398828 H. pylori GHPO 148
9	246	3.7	2893	19	AA371556 Helicobacter poly
10	241.5	3.6	1315	20	AA308642 S. aureus SdrD pro
11	239	3.6	2902	22	AA36351 H. pylori HPN165 p

12	222.5	3.4	1844	21	AA18250	Plasmodium falcipa
13	222	3.3	2366	17	AA395011	C. difficile toxin
14	222	3.3	2365	19	AA368388	Clostridium diffi
15	220	3.3	5795	22	AA37017	Staphylococcus aur
16	219	3.3	2086	22	AA34143	Staphylococcus aur
17	212	3.2	6281	22	AA37403	Staphylococcus aur
18	211	3.2	1778	22	AA352677	Escherichia coli p
19	209	3.1	2710	19	AA35016	C. difficile toxin
20	209	3.1	2710	17	AA368387	Clostridium diffi
21	208	3.1	2353	17	AA39393	Haemophilus adhesi
22	206	3.1	1222	21	AA301830	H. Influenzae stra
23	206	3.1	1228	21	AA301828	Haemophilus influe
24	206	3.1	2434	22	AA34339	Staphylococcus aur
25	205.5	3.1	1221	21	AA301825	Haemophilus influe
26	205.5	3.1	1227	21	AA301824	Haemophilus influe
27	205	3.1	12295	21	AA38180	Plasmodium falcipa
28	204	3.1	1296	14	AA341198	Ct. Helicobacter
29	203	3.1	1403	15	AA34629	Mannuronan C-5-epi
30	202	3.0	1095	21	AA301835	Haemophilus influe
31	202	3.0	1101	21	AA301834	Haemophilus influe
32	201	3.0	2053	22	AA369135	M. catarrhalis str
33	200.5	3.0	1938	17	AA309255	Bacillus alkaline
34	200.5	3.0	1938	20	AA373553	Full length Pullul
35	200.5	3.0	2123	22	AA300701	Moraxella catarrha
36	200.5	3.0	2411	21	AA323860	Haemophilus influe
37	200	3.0	2500	21	AA318272	Plasmodium falcipa
38	199.5	3.0	1802	21	AA383170	Cell wall protein
39	199.5	3.0	1802	21	AA370119	Staph. epidermidis
40	199	3.0	5024	22	AA382935	S. epidermidis ope
41	197.5	3.0	1475	11	AA308221	Recombinant alpha
42	197.5	3.0	1612	19	AA365088	R. prowazekii S-la
43	196	3.0	1144	16	AA376059	Mycoplasma pirum a
44	195	2.9	1541	11	AA307304	IgA1 protease. Ha
45	194.5	2.9	1092	19	AA341602	Staphylococcus epi
46	194.5	2.9	1155	22	AA38343	S. epidermidis ope
47	194.5	2.9	1601	18	AA330292	Non-typeable Haemo
48	193.5	2.9	1529	14	AA341732	High molecular wei
49	191.5	2.9	1048	22	AA37490	Staphylococcus aur
50	191	2.9	1180	21	AA301845	Haemophilus influe
51	191	2.9	1188	21	AA301844	Haemophilus influe
52	191	2.9	1726	21	AA357572	Moraxella catarrha
53	191	2.9	2032	20	AA300238	Enterococcus faeca
54	191	2.9	2032	20	AA300240	Enterococcus faeca
55	191	2.9	2032	20	AA300242	Enterococcus faeca
56	190	2.9	1029	22	AA343389	Staphylococcus aur
57	190	2.9	1287	16	AA379944	Helicobacter pylor
58	189.5	2.9	630	18	AA326603	Alpha-amyase-CenA
59	189.5	2.9	630	18	AA32523	Alpha-amyase-CenA
60	189.5	2.9	782	19	AA357436	Terminyl-linker-C8
61	189.5	2.9	2383	21	AA315945	E. coli proliferat
62	189	2.8	1974	19	AA398391	H. pylori GHPO 57
63	187.5	2.8	516	18	AA311326	Alkaline liquefyn
64	187.5	2.8	750	18	AA323600	Alpha-amyase-cell
65	187.5	2.8	750	18	AA323600	Alpha-amyase-cell
66	187.5	2.8	862	20	AA387371	Toxin A immunogeni
67	187.5	2.8	1104	21	AA323856	Haemophilus influe
68	187.5	2.8	1104	21	AA323859	Haemophilus influe
69	187.5	2.8	1569	22	AA398842	E. coli growth and
70	186.5	2.8	653	19	AA357437	NAM1.0 fusion cons
71	186	2.8	1288	18	AA355347	H. pylori ORF 14ee
72	186	2.8	1288	18	AA355347	H. pylori ORF 14ee
73	185.5	2.8	1337	21	AA330504	A calcium-dependen
74	185.5	2.8	1516	21	AA381895	Plasmodium falcipa
75	185	2.8	1316	21	AA330505	A calcium-dependen
76	185	2.8	1881	21	AA344506	Streptococcus pneu
77	184.5	2.8	487	11	AA309359	Sequence of amylas
78	184.5	2.8	516	21	AA35714	Mutant alpha-amyla
79	184.5	2.8	1536	14	AA341725	High molecular wei
80	184.5	2.8	3158	22	AA370118	Staphylococcus aur
81	184	2.8	482	18	AA308196	B.licheniformis mu
82	184	2.8	1962	12	AA310558	Mutant protease (A
83	184	2.8	2285	20	AA398149	Bacillus subtilis
84	183.5	2.8	483	15	AA358726	M15R alpha-amyase

85 183.5 2.8 1536 14 AAR41723 High molecular wei
86 183.5 2.8 1536 18 AAR30293 Non-typeable Haemo
87 183.5 2.8 1959 12 AAR10562 Mutant protease (d
88 183.5 2.8 1962 12 AAR10560 Mutant protease (K
89 183.5 2.8 1962 12 AAR10561 Mutant protease (N
90 183.5 2.8 1962 12 AAR10557 Mutant protease (A
91 183.5 2.8 1962 12 AAR10559 Mutant protease (A
92 183.5 2.8 1968 12 AAR10941 Mutant protease (d
93 183.5 2.8 1974 12 AAR10940 Mutant protease (d
94 183 2.8 574 19 AAW57434 Termamyl-linker-CB
95 182 2.7 482 18 AAW08199 B.licheniformis mu
96 182 2.7 501 21 AAB14822 Bacillus sp. lique
97 182 2.7 1536 15 AAR63505 Haemophilus high m
98 182 2.7 1536 21 AAB01846 Haemophilus influe
99 182 2.7 2514 21 AAY75097 Neisseria meningit
100 181.5 2.7 483 14 AAR31362 B.licheniformis al

ALIGNMENTS

RESULT

1
AAB10667

ID AAB10667 standard; Protein; 2057 AA.

XX AC AAB10667;

19-JAN-2001 (first entry)

XX DT

DE DE

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

SQ Sequence 2057 AA:

Query Match 40.5%; Score 2690; DB 21; Length 2057;

Best Local Similarity 45.7%; Pred. No. 4.5e-154;

Matches 604; Conservative 175; Mismatches 385; Indels 158; Gaps 40;

QY 30 LTNDQYCALNGQLVLLRFSKAADGNPNKSGNTVTDQFSKNYATYGGNFYVVKVNGNOVEF 89
DB 153 itgghyenhngyfyvyaasqkvgtglgnldgnl-qyfdangyqvkgysfr--dvngkhiyf 209
QY 90 SCMHATQCSNDKDSQWITVIVNGK-----EVKQLVNDTKEGAAGFRNNDVYKVP 140
DB 210 dsvtgkassn-----vdvngkaggydaggnqkksyvdassggtyyfdng-----qp 258
QY 141 ALESSM-----SQFGIILPVYVKNENVQLVHREFSNDVKTGE----- 179
DB 259 lig-qtidgnlqyfnqggvgkigggf-----dvndkriyafapntgnavanteilingkl 311
QY 180 -----GNVDFWSELMPVKDSFOK-----GNGLKQFGLQINGQYQYIDPTTG 223
DB 312 qgrdangn-----gvknafskdvagntfyfdangvmlt-glqisgktyyld-egq 360
QY 224 QPRKNFLQSGNMWYFDSDFGTGVTNALELQFAKGTSSNQYRNGNAAYSYDDKSTENY 283
DB 361 hlrknyagtfnqgmfyfdadtgaqktaleyqdgqivsgsntentphnaaksydkssfen 420
QY 284 NGYLTDWYRKPQLKDGTTWDSKEIDMRPILWVWPNTLTQAYVINYMKQHGKLLPS 343
DB 421 dgyitadtwyrtidkngtdwtastetdmrpllmwtpdkgtganyinfmsskglgitt 480
QY 344 ALPFENADAPALNHNYSIEVQONIEKRISGTNTGNTLMLHDFVTNNPNWKNDSRVN 403
DB 481 t--ytaatsqktindaafvigaieqqlskstewlrdaidsfvktganwnkqtedea 537
QY 404 PSGIOF-QGGFLKY-ENSDLTPYANS-DYRLGLRMPIM---KQQT-YRQOEFLLANDID 456
DB 538 fdglqwgglgflayqddshrtptntdsgnrklgrqplnidgsktdtdkgsefilandid 597
QY 457 NSNPVQAEOLNWLXLLNFGTITANNDQANFDSVRVPADPNIDADIANIAQDFNAAYG 516
DB 598 nsnplvqaeglnwlhlmnfgsitgnndnauifdgrvdaevndvadllikiagdyfkalyg 657
QY 517 MD-SDAVSNKHINILEDMNHADPEYFNKIGNPQLTMDDTIK-----NSLNHGLSDATNRW- 570
DB 658 taksdanankhlsiledwnqkdpqyvnaggaqlcmadytvtqsgnsithganrsmmy 717
QY 571 -----GLDAIVHQSLADREN-----NSTENVVPIPVSVFRAHDNNSQDQIONA 613
DB 718 flidtyyngdlnkkivdknrpnsgtlvriatnsqgdkvipnysfvradydaqdpirka 777
QY 614 IRD--VTGKDYHTFTFEDEQKIDAYIODQN--SPVKKYNLYNIPASVAILLTKNTIPR 669
DB 778 midhgilkmqndtffidqlagmefykdqenpsgfkkyndynlpsayamlitnktvpr 837
QY 670 VYGGDLYTDGQYMEHQTRYDFTLNLSKRVKVVAGGSGHQTMSVG-----GNNNIITS 724
DB 838 vyygdnyleggymegktylnpvisallkarikyvggqmatdssgkdkldgetdlits 897
QY 725 VRYGKCAWATDTGDET-----RTQIGVYVNTNKLGVNDKVVVLHMGRAHKNOQVRA 780
DB 898 vrfgkmtsdqtdttdgnsqdykngdgivivgnnpdtkinnndktitlmgkahnknglyra 957
QY 781 AVLTITDGVINTSDQGAIVAMTDENGDLYLSHNLVWNGKEA---DTAVOGYANPDVS 837
DB 958 lvlsndsgidvdsddkaptlrndngdlifiktntfv--kqdgtilayemkgsinalis 1015
QY 838 GYLAVVWPVGCASDNODARTAPSTEKNSGN--SAYRFAAFDSDNVIFEAFSNFYVPTTKE 895
DB 1016 gylgvvwpvgasdsqdarlv-atesssngdsvfhnaaldsnvlyegfsnfqamptspe 1074
QY 896 ERANVRIAQNADEFASLGTSFPMAPQYNSKD-----RTFLDSTIDNGYAFTRDYDLGM 950

QY 1109 GEKKQNEGFEV-KNDGNTYFYDLAGNWKNTFIEDSVGNWYFFDQDGMVENKHFVDVD 1167
 Db 1167 ---eifgsgfrvvgddvq--lsiggylakntfiqvgabqwyfdkngmvtgeqvd-- 1219
 QY 1168 SYCEKGYTFPLKNGVFRGL-----VQTDNGTY-----YFDNYG 1202
 Db 1220 --gkk--yflndglqlrhvrgsgsgghvyvypkvgafngfydfagprqdvryfdng 1275
 QY 1203 KVMRNQTINAGAMIYTLDENG 1223
 Db 1276 qmyrglhdmygttfydektg 1296

RESULT 4
 AAB26877 ID AAB26877 standard; Protein; 349 AA.
 XX AC AAB26877;
 XX DT 31-JAN-2001 (first entry)
 XX DE Glucosyltransferase C (GTF C) amino acid sequence.
 XX KW Transgenic plant; glucosyltransferase C; GTF C; glucan production;
 XX KW paper manufacture.
 XX OS Streptococcus mutans.
 XX PN US6127603-A.
 XX PD 03-OCT-2000.
 XX PF 20-JAN-1998; 98US-0009620.
 XX PR 07-JUN-1995; 95US-0485243.
 XX PA (PION-) PIONEER HI-BRED INT INC.
 XX PI Nichols SE;
 XX WPI; 2000-618381/59.
 XX N-PSDB; AAA09694.
 XX PT Transgenic plant cell for producing glucan useful in paper manufacture,
 XX PT contains DNA molecule obtained from Streptococcus mutans encoding
 XX PT glucosyltransferase C enzyme -
 XX PS Disclosure; Column 13-16; 10pp; English.
 XX CC This invention relates to a transgenic plant cell derived from maize,
 XX CC potato, cassava, sweet potato or sugarcane containing a DNA molecule
 XX CC obtained from Streptococcus mutans encoding a glucosyltransferase C
 XX CC (GTF C) enzyme (represented by the present sequence). The transgenic
 XX CC plant cells that encode GTF C are useful for producing glucan which is
 XX CC useful as a substitute for the production of modified starch and latexes
 XX CC in paper manufacture. Glucan is also useful as substitutes for
 XX CC thermoplastic molecules and imparts gloss to the paper in the coating
 XX CC step during paper manufacture.
 XX CC Sequence 349 AA;
 XX SQ

Query Match 4.9%; Score 326.5; DB 21; Length 349;
 Best Local Similarity 26.5%; Pred. NO. 5.8e-12;
 Matches 102; Conservative 49; Mismatches 139; Indels 95; Gaps 12;

QY 19 SAVSGDFTTKLTNDQYQALNGQLVLLFRFSAADG--NPSGDNVTVDQFSKNYATTCGN 76
 Db 29 tsisgslvkadstdrdqgavtesqaslvttseakettatdtststststststststst 88
 QY 77 FDYVKVNGNQVFSGHGHANQSKDSDWIIVLVNGKVKROLVNDTKGAAGFNRNDY 136
 Db 89 vs-----ttnqs-----tntanta-----nfvv 107

QY 137 KVNPALENSMSGFOGIIITLPTVTKNENVQLVHRPSNDVKTGEGNY-----DEF 186
 Db 108 kpttseqektndskliittskavn-----ltatgkfvpannntabphtvt 154
 QY 187 SELMPVKDSFGKNGPLKQ-----FGLQTINGQOYIDPTTGPQRKNFLL 231
 Db 155 dkivpikpi---glikpsslsqddiaalgcnvknirkvngkyyy-kedgtlqknyal 209
 QY 232 QSGNNWYFSDTGTGVTNALQLQFAKGTYSSE--QYRNGNAAYSDDKSIENVNGYLT 288
 Db 210 nlgkttffideigalsnntlpk--kgnitndntnsfagynqvystdvanfevhdyit 267
 QY 289 ADTWYRPKQLKDGTTWDSKETDMPILMWWPNTLTQAYLYNVMKQHGNNLLPSALPF 348
 Db 268 aeswyrykylkgdktgtstektfrpilmwtwppdetrqyynyymaqlghqt----y 323
 QY 349 NADADPAELNHYSEIVQONIEKRIS 373
 Db 324 ntatsplqlnlaeqtiqtikleekit 348

RESULT 5
 AAB69136 ID AAB69136 standard; Protein; 2314 AA.
 XX AC AAB69136;
 XX DT 24-APR-2001 (first entry)
 XX DE M. Catarrhalis lesi 200kDa protein SEQ ID NO:11.
 XX KW Moraxella catarrhalis strain Q8; major outer membrane protein;
 XX KW 200kDa outer membrane protein; antibacterial; immunogenic; infection;
 XX KW otitis media; detection.
 XX OS Moraxella catarrhalis.
 XX PN WO200107619-A1.
 XX PD 01-FEB-2001.
 XX PF 26-JUL-2000; 2000WO-CA00870.
 XX PR 27-JUL-1999; 99US-0361619.
 XX PA (CONN-) CONNAUGHT LAB LTD.
 XX PI Loosmore SM, Sasaki K, Yang Y, Klein ME;
 XX WPI; 2001-159722/16.
 XX N-PSDB; AAF59105.
 XX PT New nucleic acid encoding Moraxella catarrhalis outer membrane protein,
 XX PT useful in protective vaccines and for diagnosis -
 XX PS Claim 1; Fig 5A-Y; 247pp; English.
 XX CC The present invention describes an isolated and purified nucleic acid (I)
 XX CC that encodes a 200 kDa outer membrane protein of Moraxella catarrhalis.
 XX CC The 200 kDa outer membrane protein (II) has antibacterial activity and
 XX CC can be used in vaccines. (II), and its truncated versions, are used as
 XX CC immunogenic compositions and vaccines to protect against M. catarrhalis
 XX CC infections, particularly otitis media in humans. (II) is also used as
 XX CC antigen in immunoassays for detecting specific antibodies (Ab), and to
 XX CC generate Ab. (I) are used for recombinant production of (II) and its
 XX CC fragments are used as probes for identifying/cloning 200 kDa protein
 XX CC genes from other strains, and for diagnostic detection of M. catarrhalis.
 XX CC (I) makes possible production of large amount of recombinant immunogens.
 XX CC Expression of truncated versions of (II) reduces toxicity of the protein
 XX CC towards the Escherichia coli host. The present sequence represents the
 XX CC M. catarrhalis lesi 200kDa protein, which is given in the exemplification
 XX CC of the present invention.

The invention relates to antisense inhibitors of genes essential to prokaryotic cellular proliferation, their use in identifying the genes, their use in the discovery of novel antibiotics, the essential genes themselves and the encoded proteins. The prokaryotes used are *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The invention is also useful for the identification of potential new targets for antibiotic development. The antisense nucleic acids can also be used to identify proteins used in proliferation, to express these proteins, and to obtain antibodies capable of binding to the expressed proteins. The proteins can be used to screen compounds in rational drug discovery programmes. The antisense nucleic acid sequence is also useful to screen for homologous nucleic acids which are required for cell proliferation in a wide variety of organisms. The present sequence represents an essential prokaryotic cellular proliferation protein.
 Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 Query Match 3.7%; Score 249; DB 22; Length 1349;
 Best Local Similarity 20.6%; Pred. No. 1.8e-06;
 Matches 254; Conservative 140; Mismatches 403; Indels 434; Gaps 64;

257 KGTYS--NEQYRNGAAVSVDKSIENVNG-----YLTADTYRPRKQILKD-GTTWTD 307
 93 kemsagqnettsngkl--lekesvqttgnkverstaksdeqspkstnedlntkqti 150
 308 SKETDMRPLMWWPNTLTQAYLYNMQHGMLLPSALPFFNADA--DPAELNHYSEIVQ 365
 151 snqealqp-----dlqensvuvvqptneenkvdaktesttlnvksaak 196
 366 QNTERISGTGTDWLTLMHDFVNTNPMWKNDSRNVSFGIOFGGFLKYNSD-LRPY 424
 197 snde-----tlvdnns-nnnen-----nadiilpk 221
 425 ANSDYRLGRMPT-----NKKQTYRQGFLLANDIDNSNPVVQAEQLNWL--- 470
 222 stapklrntmrilaavqspsteakvnd-lisnttlvtvdadknkkipaqdyislksq 280
 471 -----YYLL-----NFGTITANNQANFDSVRVQADPHI--- 499
 281 itvdkvksdyftikysdvvgvlglnpdiknigdkpnngetiatatakhtannlity 340
 500 ----DADLNTAQDYFNAAYGNDSDAVS-NKHINILEDWNHADPEYFNIGN----- 546
 341 tftdyvdrfnsvqmglnyslymdadtipskn-----dvfnvigtntltkttan 390
 547 ---PQLTMDDTKNSLNHGLSDATNRWGLDAIVHQSADREN-----NSTENVVI 593
 391 iqypdyvne-knsigsaftevtshvg-----dkenpgyvkqtiyvpnsensit 438
 594 PNTSFVRAHDNNSQDQIQNAIRDVT-----GKDYHTFTEDEKGDIDAYI 638
 439 naklkvgayhssypnnglqinkdvtidkiyqvpkgtlnkgydntkeltv---tnqyl 495
 639 Q-----DONSTVKKNLYNTPASAYILLTNKQDTPRVYGDLYTQGGYMEHOTRYDYL 693
 496 qkitygdnnseavldfg-nadsayvmvntk-----fgytnsespllvqm 538
 694 TNLKSRVKYVAGQSM---QTMVSGVGNNNILTSVRVKGAMTATDT---GTDETRQGI 747
 539 atlsatgnksvsgnalgftrngsgaggev-----ykigvyywedtnkngvqelgekvy 593
 748 G-----VYVSNTPNKLKLG--VNDKYYVLLHGAHNNKQYRAAVLTTTDCGVINYSQQGAPVA 801
 594 gnvttvtdntntkgeavtkedgslipnlpngdyrvfvefnlpgkgyevtpskqgnnee 653
 802 MTDENGDLYLSSHMLVNVNGREE-----ADTAVQGYANPDVSGYLAV 842

654 l-dsng---lss-vitvngkdnlsadlgiykpknldcyyvwdtnknqigdqdekisgv 708
 843 WVPVGSNDQADATAPSTKNSNSNSAYRTNAAFDSDVIFAFSNFVYTPPKESERANVRI 902
 709 tvtl-kdengdviktvttd---adgkykftdlhngnykvefttpegtyptvtvtsgsdiek 764
 903 AQNADFFASLGTSFENAPQYNSKORTFELDSIDNGYAFTRDYDLG---MSEPKNKYGTD 959
 765 dsn-----glttgvi-----ngad-----nmldsfytkpynlgnvwdtnkdqkg 809
 960 EDLRANAQLAKAGLQVMAWVPDQIYNLFCKEVATVTRVDRGNVWKAIIINNLYVWN 1019
 810 dsteqgisgv-----tvllknengev-----lqttk 835
 1020 TIGGEYQ--KKYGGAFDLKQLYPIFFIKYKQVSTGV--AIDPSQITWSAKYFNGTN 1075
 836 tdkdgkyftglengty--kvefetsgytptqvgsgtdegids-----ngts 881
 1076 ILHRSGSVVLK-----ADGGOY---YNLG-----TTTKQFLPIQLTGEK-----K 1112
 882 -----tgvikdkdndtidsgfykptynlgyvwdtnkn--gvqdkdegisgvvtlk 934
 1113 QGNEGPFVK---GNDGNYFYDLAGNMYKNTF----- 1140
 935 dendkviktvttdengkyqftdlngtykvefetsgytptsvtsngndeksdnglittg 994
 1141 -IED-----SVGN--WFFPDQDKMYENK--HFVDVDSYGEKGYTFEL 1178
 995 vikdadmtldsgfyktpkyslgdyvvydsnkdgkqdstekgikdvkvtlinekgevi-- 1052
 1179 KNGVSFRGLVQTD-NGTYYPDNY--GK-----MVRNQT-----INAGAMIY 1217
 1053 -----gtkttdengkykcfanldsgkyvifekpagltqvtvntteddkdaggdevdv 1104
 1218 TLD-----ENGLIKASYNSDAEPYPTSTD 1241
 1105 titdhddftldngyfeedtsdsdsdsd 1135
 RESULT 7
 AAU37544
 ID AAU37544 standard; Protein; 1349 AA.
 XX AAU37544;
 XX
 DT 14-FEB-2002 (first entry)
 XX
 DE Staphylococcus aureus cellular proliferation protein #1714.
 KW Antisense; prokaryotic cellular proliferation protein;
 KW antibiotic; antibacterial; drug design.
 XX
 OS Staphylococcus aureus.
 PN WO2001709355-A2.
 PD 27-SEP-2001.
 PF 21-MAR-2001; 2001WO-US09180.
 XX 21-MAR-2000; 2000US-191078P.
 PR 23-MAY-2000; 2000US-206848P.
 PR 26-MAY-2000; 2000US-207727P.
 PR 23-OCT-2000; 2000US-242578P.
 PR 27-NOV-2000; 2000US-253625P.
 PR 22-DEC-2000; 2000US-257931P.
 PR 16-FEB-2001; 2001US-269308P.
 XX (ELIT-) ELITRA PHARM INC.
 PA
 XX
 PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
 Yamamoto RT, Xu HH;

XX WPI; 2001-611495/70.
 DR N-PSDB; AAS55403.
 XX
 PT New polynucleotides for the identification and development of
 XX antibiotics, comprise sequences of antisense nucleic acids -
 PS
 XX Example 3; Seq ID No 13137; 511bp; English.
 XX
 CC The invention relates to antisense inhibitors of genes essential to
 CC prokaryotic cellular proliferation, their use in identifying the
 CC genes, their use in the discovery of novel antibiotics, the essential
 CC genes themselves and the encoded proteins. The prokaryotes used are
 CC *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella*
 CC *pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The
 CC invention is also useful for the identification of potential new targets
 CC for antibiotic development. The antisense nucleic acids can also be used
 CC to identify proteins used in proliferation, to express these proteins,
 CC and to obtain antibodies capable of binding to the expressed proteins.
 CC The proteins can be used to screen compounds in rational drug discovery
 CC programmes. The antisense nucleic acid sequence is also useful to screen
 CC for homologous nucleic acids which are required for cell proliferation in
 CC a wide variety of organisms. The present sequence represents an
 CC essential prokaryotic cellular proliferation protein.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 1349 AA;

Query Match 3.7%; Score 249; DB 22; Length 1349;
 Best Local Similarity 20.6%; Pred. No. 3.8e-06;
 Matches 254; Conservative 140; Mismatches 403; Indels 434; Gaps 64;

QY 257 KGTYS--NEQYRNGNAVSDDKSTENYNG-----YLTATWRPKOILKD-GTWTWD 307
 DB 93 kemvsqgnettsgnkl-----lekesvqstgtnkvevstaksdeqaspkstnedlnkqti 150
 QY 308 SKETMRPILVMWPNLTQAYLVNFMKORGLPSALPFFNADA--DPAELNHYSEIVQ 365
 DB 151 snqea.lq-----dlqenksvvnvqptneenkvdaktesttlnvksdalk 196
 QY 366 QNTEKRISGETNFDMLRLMHDVPTNPMWKNKDSNNVFSIQFGGFLKYENS-DLTPY 424
 DB 197 snde-----tlvdhns--nsnne-----nadiilpk 221
 QY 425 ANSDYRLGRMPI-----NKKDQYRGQEFLLANDIDNSNPVVOAQLNML--- 470
 DB 222 stapkrlntmrifaavqssteakvnd-litstltvtvdadknnkiypaqdyislksq 280
 QY 471 -----YYLL-----NFGPITANNOQANFDSVRVDAPDNI--- 499
 DB 281 itvdkvksgdyftikysdvtvqygnpedlknlgdipnnggetiatakhdtannlity 340
 QY 500 ----DADLNIADQYFNAAVCHMSDAVS-NKHNIILEDWHADEYFNKIGN----- 546
 DB 341 tftdyvdrfnsvqmgnginsymdadtipskn-----dvefnvgtntttkttan 390
 QY 547 ---POLNMDDTIKNSLNHGLSDATNRWGLDAIVHQSLADREN-----NSTENVVI 593
 DB 391 iqydpdyvne--knsigsafetvshvg-----nkenpgyykqtiyvnpnsensit 438
 QY 594 PNYSFVRAHONNSODQIQNAIROVY-----GKDYHTFTFEDROKGDIDAYI 638
 DB 439 naklkvaqayhssypnnigqinkvdtdikiyqvpkqytlngkydvntkeltidv---tnqyl 495
 QY 639 Q-----DONSTVKYKYNLNPASVAILLTNKDTIPRYVYGDLYTDGQYWEKOTRYIDTL 693
 DB 496 qklygdnnnsavidfg--nadsayvnmvntk-----fgytnsesptlvqm 538
 QY 694 TNLKSRVYKYVAGGQSM--QTMSVGGNNNLTLSRVYKKGAMTATDT---GTDERTQGI 747

DB 539 atlsstgnksvstgnalgftnnqsgagqev-----ykginyvwdtnkngvqelgekgv 593
 QY 748 G-----VYVSNTPLKLG--VNDKVLHMGAAHKNQOYRAAVLTTTGGVINYTSDQCAPVA 801
 DB 594 gnvvtvfdntntkvgavtkedsgsylvlnpnpngdyrvfnsnlpkgayevipsgkgnnee 653
 QY 802 MTDENGDLXLSHNLVYNGKEE-----ADTAVOGYANPDVSGYLAV 842
 DB 654 I-dsng---lss-vitvngkdnlsadlgiykpkynlodyvwdtnknglqddkdegisgv 708
 QY 843 WVPVGSADNDQARTAPSTKNSGNSAYRTNAAFDSNVIFEAFSNEVYTPPTKESRANVRI 902
 DB 709 tvti-kaengdvktvtd--adgkykfddlhngnykveftipegytptvtvsgsdelek 764
 QY 903 AONADFPASLGFTSFEMAPQVNSKDRFTFDSIDNGYAFTRDYDLG---MSEPNKYCTD 959
 DB 765 dsn-----gltttgvl-----ngad-----nmtidsfytkpynlgnvymwedtnkdqk 809
 QY 960 EDLRNATQALHKAGLOYMADWVPDIYNLPKGKAVATVRVDRGNVWKDAIINNLYVWN 1019
 DB 810 dstekgsgv-----lkl-----vtvlknengev-----lqtk 835
 QY 1020 TIGGGEYQ--KKYGAFLDKLQKLYPEFTKKOYSTGV--AIDFSOKITEWSAKYFNGTN 1075
 DB 836 tdkgkyqftglengty--kvefetpsgytptqvgsgtdegids-----ngts 881
 QY 1076 ILHRGSGVVLK-----ADGGQY---YNLG-----TTKQFLPIQLGKEX-----x 1112
 DB 882 -----ttgvilkdkndtidsgfykptynlgyvwdtnkn--gvqdkdegisgvvtvltk 934
 QY 1113 QGNEGFVK---GNDGNYFYFDLAGNVMVKNTF----- 1140
 DB 935 dendkvltvtdengkyqfcdlnngtykvefetpsgytptsvtsngntekdsngltttg 994
 QY 1141 -TED-----SVGN--WYFFDQDGKMWENK---HFVDVDSYGEKGTFFFL 1178
 DB 995 vlkdadnmtldsgfyktpkyslgdyvwydsnkdgkqdatekgikdvkvtllnekevl-- 1052
 QY 1179 KNGVSFRGLQVQTO-NGVYFDNY--GR-----MVRNOT-----INAGAMIY 1217
 DB 1053 -----gtktdekgkyfcdlnldsgkykvefepagltqvtvttteddkdadggevdv 1104
 QY 1218 TLD-----ENGLIKASYNSDAEYPTSTD 1241
 DB 1105 titdhddfldngyfeedtsdsdsdsdsd 1135

RESULT 8
 AAW98828
 ID AAW98828 standard; Protein; 2893 AA.
 XX
 AC AAW98828;
 XX
 DT 31-MAR-1999 (first entry)
 XX
 DE H. pylori GHPO 1484 protein.
 XX
 KW GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;
 KW peptic ulcer disease.
 XX
 OS Helicobacter pylori.
 XX
 PN WO9843478-A1.
 XX
 XX 08-OCT-1998.
 XX
 PF 01-APR-1998; 98WO-US06371.
 XX
 PR 29-JUL-1997; 97US-0902615.
 PR 01-APR-1997; 97US-0833457.
 PR 24-JUN-1997; 97US-0881227.
 XX

PA (HUMA-) HUMAN GENOME SCI INC.
 PA (INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.

PI Al-Garawi A, Kleanthous H, Miller C, Oomen RP, Tomb J;
 XX N-PSDB; AAX14547.

XX WPI: 1998-542293/46.

DR N-PSDB; AAX14547.

XX New isolated Helicobacter polynucleotides - used to develop products
 PT for the diagnosis, prevention and treatment of Helicobacter
 PT infections and gastrointestinal diseases

XX Claim 8: Page 1827-1840; 2054pp; English.

XX This sequence represents a Helicobacter pylori GHPO protein of the
 CC invention. The polypeptides can be used for preventing or treating
 CC Helicobacter infections, and gastrointestinal diseases associated with
 CC these infections, including acute, chronic, and atrophic gastritis, and
 CC peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be
 CC used for the production of antibodies. The products can also be used for
 CC detection and diagnosis.

XX Sequence 2893 AA;

Query Match 3.7%; Score 246; DB 19; Length 2893;

Best Local Similarity 19.4%; Pred. No. 7.9e-06;

Matches 285; Conservative 181; Mismatches 571; Indels 434; Gaps 71;

QY 7 PDVAAYSLYNSAVSGDPTIKLNDYQALNGLOVLLRFSAADGNPSGDWVTPQF 66

DB 353 pnnnaghpqignsfsgnattikgfvnfgqafnns-nhqltqgnasfnfnatnngkkti 411

QY 67 SKNYATTGCFYVNVNGQVFEFGSHATNOSNDKDSQWIIIVLWNGKEVKROLVNDTKEG 126

DB 412 ekdasfnntfn-tsvdtnnmsvtgg-----vltsgk-----ndllkag 448

QY 127 AA-GFNRNDVYKVPAINSSMSGFGIITL-----PVTVKENYVQLVHVFNDVKTGEG 180

DB 449 stldfssktilagttfnltsgkskvstlinsggitysnllnhaingltsalktnes 508

QY 181 NVDFWSELPVDFSPQK-----NGPLKFGIQTNGQYQYIDPTGQPKNFLLQS 233

DB 509 -----isnpqsfagldlityngvtgql-----lnenaatsktpdspsks-----s 551

QY 234 GNNWIFDSDTCVGTNALQLPAKGVSSNEQYRNGNAAYSDDKSIEN-----VNGY 286

DB 552 tntstqvvygkigtikyklq--elfshnsliiqaesgtytppvingskfdlsasny 608

QY 287 LTAD-TWYRPKQILKDGTTWOSKETDMRPILMVWPNTLQAYILNYMKQGNLPSAL 345

DB 609 inadmwydhkyypksqntes-gtyilpsvqiw-----gsyntsfk----- 651

QY 346 PFNRADPAELNHYSEIVQONIEKR-----ISEGTNDWLRT 383

DB 652 -tfisangslvlygnstwdhnmvsssgtvsfgdtsalngchcpwpyyqctgthgty 710

QY 384 LMHDFVT-----NNPMNN-----RD 398

DB 711 ayhvyitanlrsgnrigtggaaanlfnvgvdsinianatitqhnaglyssmsftsgmdn 770

QY 399 SENNV-----FSGIQFGGLKYNSDLTPYAN 426

DB 771 sqnlinginsngkisyvgttftneakgkfnfnagqavfentnngsyqfsgdsln-fsn 829

QY 427 SDYALLGRPNINKQTYRQEF-----LLANDIDSNPVQ-----A 464

DB 830 mngfngsgfelsaknasfnanfnnsasfnnsnattsfvfgdfnansnlqagnavfg 889

QY 465 EQLNMLYLLNFGTITANN--DQANFDSVRVADPNTI-----DADLMNIAQDYFNA--AYG 516

DB 890 nstngsqntanfnntgsvnlsnagnatfnvfnvngptntsvkqvtlnnltiklnaplsfg 949

QY 517 MDSDAVSNKH--INILEDNRHADPEYFNKIGNPOLTMDPTIKNSLNHGLSDATNRHGL-- 572

DB 950 -dgtitfnahsviniaesitn-----gnptivsskskeieynnafs--knlwqlin 997

QY 573 ---DAIVHQSLADRENNSTENVIPNYSFVRAHDNSQDQIQNAIRDVTKDYHTTFED 629

DB 998 yqhgassekvlsvsagngvydvv---ysfnqntynfgevfsqnsisirlrgvmvdyvd 1054

QY 630 EOKGIDAYIQDONSTVKKYNYLNIPASYAILLANKDPIPRVYGDLYTD---GGQYMEHQ 686

DB 1055 meksdhlyygn---algfnty-mpnsynnnlgnann--tiyyvksidfyasgkltfk 1107

QY 687 TRYDYDTILNLLSKSVKYVAGOSMOT-MSVGGNNNLTISVRYG--KGAMTATDTGDETR 743

DB 1108 aetsqtfq---qnsaivfgakswtsidsdpqsn--tiirfgnkgg--gsndasghwn 1161

QY 744 TQIGGVVSTPNLKLGVNDKV-----VLHMGAAHKKQYRAAVALTTTDDGVINYTSQGA 798

DB 1162 lqclgfitghyeakkiyitgsiesgnrissggaslnfnglqlltnatyn----- 1214

QY 799 PVAMTDENDLYLS-SHNLVYNGKERADTAVQGYANPDVS-----GYLAVW 843

DB 1215 raagtqssmufisnsanigqnsyfidtdaqngnbnfnalnldfsnssfrgy--- 1270

QY 844 VPVGASDNQ---DARTAPS--TEKNSGNSAYRTNAAPDSNVIPFAFSNFTPTKESERA 898

DB 1271 --vgktgevfknaknalsfnstnslsglygmqa---ksvifdn-snlsvsgtssika 1324

QY 899 N-VRTAQNAADFFASIGTSTFEMAFQYNSSKDRTELD---STID--NGYAFTRDYDLQWSE 952

DB 1325 nainlsqnasinas-nhstlelqgdlwn-dtssalnqstlnvsnnatindyasllasn 1382

QY 953 PNKYGTDDRLRNAQALHAKAGLOVMAHWVPQIYNLPGEKVAFTVRVDDRGVWAKDAIIN 1012

DB 1383 ---gshlnfnfgavn-fnsani-----ttslnnssivfkavsl 1416

QY 1013 NNLVYVNTIGGEYQKKGGAFL--DKLQKLYPEITFKQVSTGVADIPDSOKITEWSAKY 1070

DB 1417 g9qfnlmsnslldf---gssaitentafnfydnafsqspitfhqaldikaps----- 1467

QY 1071 NRGNTNLRHSGSYVLKADGGYNYNGTGTQKFLPQLTGERKKQGNFVKQNDQ--NYVF 1128

DB 1468 -lqgnllpnssvldlknsg-----lvfdqggsalnian 1500

QY 1129 YDLAG-----NMVKNFTIEDSVGNWYFDDQDKAVENKHFVDVDSYGEKGTYPFLKNGV 1182

DB 1501 idllsdlndknrvynliqadmnsnw-----erisff-----gm 1535

QY 1183 SFRGLVQTDNGTYTF-----DNYGKM-----VRNOTINAGAMIYTL 1220

DB 1536 hindglydakngctyftnplnallkitesfkndqlsvtlisqilpgikntlynlgsseifnyq 1595

QY 1221 E-----NGKLIRKASYNDAE---YPTSTDVG 1243

DB 1596 kvynnangvy---sysddaqgvfyitsnvkg 1623

RESULT 9
 AAW71556
 ID AAW71556 standard; Protein; 2893 AA.
 XX AAW71556;
 AC AAW71556;
 DT 09-NOV-1998 (first entry)
 XX Helicobacter polypeptide GHPO 1484.
 DE GHPO 1484; infection; therapy; diagnosis; vaccine; gastritis;
 KW ulcer.
 XX Helicobacter pylori.
 OS Helicobacter pylori.
 XX WO9821225-A1.

22-MAY-1998. 14-NOV-1997; 97WO-US21353. 29-JUL-1997; 97US-0902615. 14-NOV-1996; 96US-0749051. 01-APR-1997; 97US-0831309. 01-APR-1997; 97US-0833457. 01-APR-1997; 97US-0834705. 24-JUN-1997; 97US-0881227. (HUMA-) HUMAN GENOME SCI INC. (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN. (INNR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS. Al-Garawi A, Haas R, Kleanthous H, Meyer T, Miller C; Odenbreit S, Tomb J; WPI: 1998-297855/26. N-PSDB; AAV52091. Helicobacter polynucleotide and polypeptide sequences - useful to treat or prevent gastrointestinal infection Claim 1; Page 330-337; 362pp; English. This claimed Helicobacter pylori polypeptide, designated GHPO 1484, can be used in vaccination methods for preventing or treating Helicobacter infection. 85 Helicobacter polypeptides (see AAW1474-W1558) are claimed, as well as isolated polynucleotides (see AAV52009-93) that encode them. The invention also provides: methods for producing these Helicobacter polypeptides in recombinant host systems, and related expression cassettes, vectors and transformed or transfected host cells; live vaccine vectors that contain the polynucleotides of the invention and which can be used to prevent or treat Helicobacter infection; therapeutic and/or prophylactic methods involving administration of polynucleotide molecules, polypeptides or monospecific antibodies; methods for detecting the presence of Helicobacter in samples using e.g. the polypeptides or monospecific antibodies; and methods for purifying the polypeptides by antibody-based affinity chromatography. Sequence 2893 AA; Query Match 3.7%; Score 246; DB 19; Length 2893; Best Local Similarity 19.4%; Pred. No. 7.9e-06; Matches 285; Conservative 181; Mismatches 571; Indels 434; Gaps 71; QY 7 PDVAWVPSLYNSAVSGFDTTILKTDQYQALNGQLVLLRFSSKAADGNPSGDNTVDOF 66 Db 353 ptnaahpqhqnasfsgnatllkgvfnfqafnns-nhqltiqnasfnnatfnatgkiti 411 QY 67 SKNYATGGFDFVKKVNGNOVERSGHHATNQSDKDSQWLIYLVNGKEVKRQLYNDTKG 126 Db 412 ekdasfnntfn-tsvdtnmsvtgg-----vtlsgk-----ndlknq 448 QY 127 AA-GFNENDVKKYNPALENSMSGFGIITL-----PVTVKENVQLVHRESNDVKTGG 180 Db 449 stldfsgskktllaogtftnitslgseksvtlmsggitysnlhlnhaigltsaalktnes 508 QY 181 NYVDFWSELMPVKDSFOG-----NGPLKQFGLQTINGQQYYIDFTTGPQPKNFFLOS 233 Db 509 -----lsnpqsfagqlwdlityngvtgql-----lnenaatskptdspsks-----s 551 QY 234 GNWIIYFDSGTGVTNALELQFAKGVSSNEQYENGAAVSYDDKSLEN-----VNGY 286 Db 552 tnstgvqvgykgidgtiykq---etfshnsiilaqlesgtytpppvgingskfdlsasny 608 QY 287 LTAD-TWYRPKQILKDGTTWDSKETDMRPILMWVWNTLTQAYLYNKKQHGNNLLPSAL 345

QY 1221 E-----NGKLIKASYNSDAE---YPtSTDVG 1243
:
DB 1596 kvynnangvv--svsdgaqgvfyltsnvkq 1623

RESULT 10
 AAY08642
 ID AAY08642 standard: Protein: 1315 AA.

AA	AA
AC	AA08642;
XX	
DT	09-AUG-1999 (first entry)

DE *S. aureus* SdrD protein.

KW KW Fibrinogen-binding protein; alpha chain; beta chain; ClfB; SdrC; SdrD; SdrE; fibrinogen; medical device; competitive inhibitor; pharmaceutical; treatment; infection; septicemia; osteomyelitis; mastitis; endocarditis; extracellular matrix; vascular graft; vascular stent; vaccine; intravenous catheter; artificial heart valve; cardiac assist device; antibacterial. KW

05 staphylococcus aureus.

PN W09927109-A2.

03-JUN-1999.

25-NOV-1998: 98WO-US25246.

XX
PR 37-AUG-1998: 98US-0098427.

PR 26-NOV-1997; 9705-0000013;

PA (EIDH/) EIDHIN D N.
PA (EIDE-) EIDESE T/A BIORESEARCH IRELAND

PA (FOST) FOSTER T J.
 PR (HOOK) HOOK M A C

PA (INHI-) INHIBITEX I

PA (PATT/) PATTI J M.

XX

PI Perkins SE;

DR WPI; 1999-357844/30

XX

PT
septicemia, osteomy

PS Claim 8; Fig 8; 143pp; English.

This invention describes novel staphylococcus aureus fibrinogen-binding proteins that bind both the alpha and beta fibrinogen chains. The proteins (and their encoding nucleic acids are ClfB, SdrC, SdrD and SdrE). Staphylococcus aureus is thought to utilize fibrinogen to adhere to medical devices, binding proteins that bind both the alpha and beta fibrinogen chains (ClfB, SdrC, SdrD and SdrE) can therefore be used as competitive inhibitors to block this binding. Antibodies against ClfB, SdrC, SdrD and SdrE inhibit ClfB, SdrC, SdrD and SdrE mediated binding. The proteins of the invention can be used in a pharmaceutical composition for the treatment of Staphylococcus aureus infection e.g. septicemia, osteomyelitis, mastitis or endocarditis or to inhibit the binding of S. aureus to the extracellular matrix. The proteins or their fragments may be used to coat a medical device to reduce the S. aureus infection of an indwelling medical device, especially where the medical device is selected from the group consisting of vascular grafts, vascular stents, intravenous catheters, artificial heart valves, and cardiac assist devices. ClfB, SdrC, SdrD, SdrE, or an active fragment, subdomain or encoding gene may be used as a vaccine. The DS (aspartate serine) repeat region or a gene encoding it may be used as an identifying probe for the

Db 831 -----lqtktdkdgkyqftglengty--kvefetpsgytptqvgsgtdgids----- 877
 QY 1066 WSAKYFNGTNIHRSGYVLK-----ADGGQY---YNLG-----TTTKQFLPIQLTGEX 1111
 Db 878 -----ngts-----ttgvikdkdndidsqfykptynlgydwedtnkn--gvqdkdek 924
 QY 1112 -----KOGNEGFVK-----GNDGNYFYDLAGNNVKVTF----- 1140
 Db 925 gisgvtvilkdndkvikttdtdengkyqftdlnggtykvefetpsgytptsvtsngnte 984
 QY 1141 -----LED-----SVGN--WFFFDQDGKVENK---HFVDVDS 1168
 Db 985 kdsnglittgvtikdadmtdidsqfyktpkysiggyvvydsnkdkqdstekgikdkvvtl 1044
 QY 1169 YGRKGTFFFLKNGVSFRGLVQTD--NGTYFFDN 1200
 Db 1045 lnekgevl-----gttktdengkyfcfdn 1067
 RESULT 11
 AAB46351
 ID AAB46351 standard; Protein; 2902 AA.
 XX AAB46351;
 KC AAB46351;
 DT 05-APR-2001 (first entry)
 XX H. pylori HPN165 protein.
 DE Microbial infection; antibacterial; Helicobacter pylori infection;
 KW vaccine; screening.
 XX Helicobacter pylori.
 OS WO200073502-A2.
 PN 07-DEC-2000.
 PD 31-MAY-2000; 2000WO-BP05024.
 PF 31-MAY-1999; 99DE-1024965.
 PR 17-JUN-1999; 99DE-1027740.
 PR 21-JUL-1999; 99DE-1034029.
 XX (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
 PA (CREA-) CREATOGEN GMBH.
 PI Apfel H, Fuchs TM, Gibbs CP, Hueck CJ, Meyer TF;
 XX WPI; 2001-049948/06.
 DR N-PSDB; AAF25628.
 XX Preparing an agent for diagnosis or control of microbial infection,
 PT useful particularly against Helicobacter, based on identification of
 PT essential genes in defective mutants -
 XX Claim 37; Figure 15; 366pp; German.
 PS This invention describes a novel preparation of an agent (A) for
 CC detection, prevention and/or treatment of microbial infection by:
 CC (i) identifying essential genes (I) and corresponding polypeptides
 CC (II); (ii) identifying compounds that are directed against (II) and
 CC inactivate the microbe; (iii) testing these for suitability for use; and
 CC (iv) formulating selected (A). Identifying essential genes (I) comprises
 CC preparation of gene-deficient microorganisms by conditional antisense
 CC inhibition (CAI) and/or subtractive recombination mutagenesis (SRM),
 CC then determining viability and/or survival of the deficient organisms.
 CC The products of the invention have antibacterial activity. (A) (which may
 CC be a nucleic acid (Ia), vector or host cell containing (Ia), derived
 CC polypeptide (Iia), or fragments, (Iia)-specific antibodies or their
 CC fragments or an inhibitor of (Iia)) are particularly used for diagnosis,
 CC treatment or prevention of infection by Helicobacter pylori. Particularly
 CC (Ia) and (Iia) are used in DNA, subunit or live vaccines. The method

CC identifies essential genes, including those that have homologs in other
 CC species, so identified (A) should have a broad spectrum of activity. Many
 CC gene-deficient cells can be screened quickly, in an automated process,
 CC and the identified genes can be used for screening without purification.
 XX Sequence 2902 AA:
 QY 13 YPSLYNSAVSGFTTKLTNDQYQALNGQLVLLRFSKAADGNPSGDNTVTDQFSKNYAT 72
 Db 366 hpiqnsfsgsattikgfatfeqafms-nhltlqnasfnntatnntgkkitiekdasf 424
 QY 73 TGGNFYVYKNGQVFEFSGWHATNQSDKDSQWIIIVLVNQEKVKQLVNDTKEGAA-GRN 131
 Db 425 nntsfntpvdtanmtisgg-----vltsgk-----ndlkngatldfg 461
 QY 132 RNDYKVNPAIENSNSGFGQIITL-----PVTVKVENVOLVHRESNDVKTGEGNYVDFW 186
 Db 462 sskiltqgttfnltslgseksvtilnsrggitynhllnhainstlalktne----- 514
 QY 187 SELMPVKDSFOKG-----NGPLKQFLOTINGQQYYIDFTTGPQPRKFLQSGNNWY 239
 Db 515 ssskp--qsfagglwmdityngvtgql---lnenaatskptdespsks----stnstqv 564
 QY 240 FDSDTGVTNALELOFAKGTVSSNEQYRNGNAYSYDDKSIEN-----VNGYLTAD-T 291
 Db 565 yqvdykigdtlyklyq---etfshnsliiqalesgtypppvyngskfllaasnyindmp 621
 QY 292 WYRQKQILKDGTTWTDSKETDMPILMVW----- 320
 Db 622 wynhkyypksqntfes-gtylpsvqlwgsytskfqtfsasnsnlviagnatwtchmv 680
 QY 321 -----WP-----NTLTQAY---YLYNMKQHGMLPS----- 343
 Db 681 sssdtvafgdtsgsalngchcgpwyqctgtngtyasahvyitanlrsgnrigtgaan 740
 QY 344 -----ALPFFNADAPAELENHYSEIV---QQMLEKRISGTNTDMLRLMHDFTVNNP 393
 Db 741 lifngvdsinianaaitghnagayssmtfstqumdsqnlnginsngkllvygtttfnq 800
 QY 394 MNWKDS-----ENVNPSG--IQFGGFLKYENSDLTPVANSYRLGRMPNIK 440
 Db 801 a--kdgkfifnaggtfentufngsyqfsgdslnfsnmn---qfns-----gsfelgak 850
 QY 441 DQTYRQGEF-----LLANDIDNSNPVQ-----AEQLNWLYLLNFGT 478
 Db 851 ntifnnafnmstsfnnssattsfvgdftnansnlqagnavfngstngsqntanfn 910
 QY 479 ITANN--DOANPDSYRVDAPONID-----ADLNNIAQDYFNA--AYGMDSDAVSNKH--IN 528
 Db 911 tgsuvalnatfdvfvnsptntsvkgtvlnlilklnaplsfg-dgtlvfsahsvin 969
 QY 529 ILEDNMHADPEYFNKIGNPOLTMDDTIKNSLMHGISDATNRWGL-----DAIVHQSLADR 583
 Db 970 lgealt-----gnp-iltvsssk-aielyndafsknlwqlinyqghgasseklyss 1018
 QY 584 ENNSTENVVPIPSYFVRAHDNNSOQIQ-----NAI---ROVTKDYHTFTFEDQKGID 635
 Db 1019 agngvydv---ysf-----nnqtnyfgvtfspnsislrirvgvm---vfydvmeksdr 1067
 QY 636 AIIQONSTVKYNYNINYPASVAILLTNKDTPRYVYGDLYD---GGQYEHKQTRYDIT 692
 Db 1068 lyyqn-----algfmy-mpsynnnalgnnn--tiyydmsidfysgkftlftkaefsgt 1120
 QY 693 LTNLLKSRVYKVVAGSGSMOTMSVGGNNNLTTSV-----RYG--KGAMTATDGTG 740
 Db 1121 ft-----ggp--saivfaknlwtvsdapsenviirfydnkga-gsndasgh 1165
 QY 741 ETRTQIGVYVWSNTPNLKLGVDNKV-----VLHMGAAHKNQOYRAAFLTDTGVTNYTSD 795

Db 1166 cwnlqicgfitghyeackiyitgsiesgnrissgggashnflnglqilhtnatiyn----- 1221
 QY 796 QGAPVMTDENGDLXLS-SHNLVYNGKEADTAQVYANPDVS-----GYL 840
 Db 1222 ---raqtqssmfmfvsnsanigaqnsyfidtaqngnpgnfnalnldfsnssfrgy- 1277
 QY 841 AVWVPVQASD-----NQDAKTAPESTEKNSGNSAYRTNAFDSNVIFAFNFTPTPKES 895
 Db 1278 -----vqgtqsvkfnavnalsftnssnlsqlyqmqd---ksvlfdu-snlsvsvgtas 1328
 QY 896 ERAN-VRLAQNADEFASLGTFSFMAQYN-----SSKDRFLDSTIDNGYAFTDXYDLM 950
 Db 1329 lkanalnsqnasinas-nhstlelqgdlndtsslnlqgsainvsnnatlndyaslia 1387
 QY 951 SEPKNYGVDELRNAIQALHKAGLVQVMDWPDQIYNLPKGEVATVRVDDRGVNWKDAI 1010
 Db 1388 sn-----gshlnfgavn-fnsani-----ttsissssiivfkav 1421
 QY 1011 INNLIYVNTTIGGGEYQKKYGAFI--DKLQKLYPEIPTKKQVSTGVVAIDPSQKITPWSA 1068
 Db 1422 slrgqfnlnssldfq---gsaaltsntafndafsgpifhqdldikvpls----- 1474
 QY 1069 KYFNGTNIHSGSVGLKADCGQYNYLGTGKQLPQLGTGKKGQEGFVKGNDGNVYF 1128
 Db 1475 ---lgnlnlponssvlnlksq-----lvfsdqslslnianldlslngn--- 1517
 QY 1129 YDLAGNMVKNFTIEDSVGNWYFFDQDQKVENKHVVDVDSYGERGTVPFLKNGVSRFGL 1188
 Db 1518 ---Knrnylnlqadmngwy-----erlnff-----gmrlndg1 1548
 QY 1189 VQTDNGTYF-----DNYGKM-----VRNQFINAGAMITLDE----- 1221
 Db 1549 ydakngtyftnplnnaiktesknnqlsvltsqipqikatllynigseifnyqkyuna 1608
 QY 1222 NGKLIKASVNSDAE---YPTSDVG 1243
 Db 1609 ngvy---sysddagvfytltsvkg 1630

RESULT 12
 AA18250
 ID AA18250 standard; Protein; 1844 AA.
 XX
 AC AA18250;
 XX
 DT 07-NOV-2000 (first entry)
 XX
 DE Plasmodium falciparum chromosome 2 related protein SEQ ID NO:107.
 XX
 KW Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
 KW antimalarial; malaria; protozoacide; infection; insecticide.
 XX
 OS Plasmodium falciparum.
 XX
 PN WO200025728-A2.
 XX
 PD 11-MAY-2000.
 XX
 PF 05-NOV-1999; 99WO-US26796.
 XX
 PR 05-NOV-1998; 98US-0107131.
 XX
 XX (HOFF/) HOFFMAN S.
 PA (CARU/) CARUCCI D.
 PA (GARD/) GARDNER M.
 PA (VENT/) VENTER J C.
 XX
 PI Hoffman S, Carucci D, Gardner M, Venter JC;
 XX WPI; 2000-365347/31.
 XX
 XX Proteins encoded by chromosome 2 of the human malarial parasite,
 PT

PT Plasmodium falciparum, useful as antimalarial vaccines and in the
 diagnosis of P.falciparum infection -
 XX
 PS Disclosure: Page 246-250; 577pp: English.
 XX
 CC The present invention describes proteins and their fragments (I) encoded
 by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
 CC Also described are: (I) nucleotide sequences (II) encoding (I); and (2)
 CC vaccines against P. falciparum infection comprising (I) or (II).
 CC (I) and (II) are useful for the development of vaccines against
 CC P. falciparum infection. (I) and polyclonal antisera or a monoclonal
 CC antibody raised to immunogens comprising the sequences of (I); are
 CC useful in the detection of infection with P. falciparum. Furthermore,
 CC (I) especially when they are rifins or secreted or membrane proteins)
 CC can aid the identification of drugs to treat or prevent P. falciparum
 CC infection, or they can be used to identify drug resistance in
 CC P. falciparum. Sequencing of the Plasmodium chromosome 2 and the
 CC subsequent identification of proteins encoded by it will help to expand
 CC our understanding of parasite biology, a process hampered by the
 CC complexity of the parasitic lifecycle, and provide new targets for
 CC vaccine and drug development. Parasite resistance to drugs and mosquito
 CC resistance to insecticides have led to a resurgence of malaria in many
 CC parts of the world, and there is a pressing need for vaccines and new
 CC drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352 represent nucleotide
 CC and protein sequences given in the present invention, but which are not
 CC specifically mentioned within the specification.
 XX
 SQ Sequence 1844 AA;

Query Match 3.4%; Score 222.5; DB 21; Length 1844;
 Best Local Similarity 17.9%; Pred. No. 0.00011;
 Matches 223; Conservative 182; Mismatches 501; Indels 339; Gaps 54;

QY 77 FDYKVGNGQVEFGSWIA-TNQSNDKDSQWIIIVLVNGK-EVKRLVNDTREGAAGFRND 134
 Db 10 fdeyslnnn---snlnayvmsneahnnf---lvnrndmnyemynslngshmsnlnnn 62
 QY 135 VYKYNPALENSMSGFGQIITLPVTVKNENVQLVHRSNDVKTGEGNVYDFWSELMPVKD 194
 Db 63 tnnlqdayinkeih-ymasdklnaknbqnvmtatynmdknaannmllhnhimtd 121
 QY 195 SFQKNGPLKQGLQTNGQYYIDPTTGQPKRKNFLQSG-----NNWYFSDSTGV 246
 Db 122 d-----qpyfynstnnkmntlkennfnannmtvnaasyfhtda-- 161
 QY 247 GTNALELOFAKGTVSSNEQ--YRGNRAAYSD---DKSIENVNGYLTADTWYRPKQILK 300
 Db 162 -----nflnftnqneqdyvnnnliannfnngnvdknnnnnn----- 200
 QY 301 DGTWTDSKETDMRPIIAVWVWPNTLTQAYLYNMKQHGNNLPSALPFFFNADADPAELNHY 360
 Db 201 -----pkave-----ginkfahlynaqpfbhf 223
 QY 361 SEIVQV----NTEKRISGTNDWLRTL--MHDFTVNNPMNMKNKSENVNFSGIOFGGFL 414
 Db 224 lpinisngknnglennasishnvsvstiselhnf---nymnmmdlnnvdmnmnmnmnm 280
 QY 415 KYENSGLTPYANSYLLGEMPINIKDKQTYRQCEFLANDINDNSNPVQAELNWLXLL 474
 Db 281 nm 324
 QY 475 NFGTITANN-DQANFDSVRVDAPDNID--ADLMNTAQDYFNAAYGMS--DAVSNNKHNI 529
 Db 325 nm 382
 QY 530 LEDWNHADPEYFNKIGNPQLTMDDTIKNSLNHGLSDATNRWGLDAIVHQSL----- 580
 Db 383 ithfnnglshnvvnvsn--fsnnahlndsnmmrlnavnfnfgdlndsfhlnemqvlkn 440
 QY 581 ADRENKSTENVV---IPNYSFVRAHDNNSQDIQNAIRDVYTKDYHTTFEDEQKIDAY 637
 Db 441 vamqnghfnvnmndemknyvnykrinsishipymnli-----knynehtsmvkgk----- 491

QY	853 DARTAPSTPKNSGNISAYRTNAAFDSNVIFEAFSNFVYPTTKESERANVRITQAUAFFASL	912
Ddb	I : I :	:
QY	2016 hfyfaengemqig-----vntedgkfyfahhnedlgn-----eegeeiys	2057
Ddb	I : I :	:
QY	913 GFTSFEMAPQYNSSKDRFLDSTINDIYAFTDRYL-----GMSEPNKYCTGDGLRNAIQ	967
Ddb	I : I :	:
QY	2058 gilinfn-----nkiyyfdsfavvgwkldgedskyyfdedaey-	2098
Ddb	I : I :	:
QY	968 ALHKAGLIQMADWPDIQLNLPG-KEVAVTPEVDNRGVNKDA-----IINNNLYVV	1018
Ddb	I : I :	:
QY	2099 ----lgilind--gqqyfnddgimqvgtc-indkfvyfsdsgilesgvqniddnyfi	2151
Ddb	I : I :	:
QY	1019 NTIG---GGEQKKYGGAFLDKLAQLEIFETFKQSVTGVAIDPS--QRITERSAKYFNG	1073
Ddb	I : I :	:
QY	2152 ddngivqlvgfdtsdgkykfapantvdniy-----ggaveysglivrve-dvyyfge	2203
Ddb	I : I :	:
QY	1074 TWILHRGSIVVLKADGGQQYYNLGTTRKFQFIQLTGEKKQGNEGFGVGNDGNYFYFDLAG	1133
Ddb	I : I :	:
QY	2204 tytieltgwlydmnesdkyy-----fnp-----etkkackginliddikyyif-dekg	2249
Ddb	I : I :	:
QY	1134 NMVKNTFIEDSGVMWVFDDQCKM-----VENKHFDVDSDSYEGKGTYFFELKNGVSPRG	1187
Ddb	I : I :	:
QY	2250 --imrtgllisfcennynynnemgmfgviniedkmfy----fgcdgv-----mqig	2294
Ddb	I : I :	:
QY	1188 LVQTDMGTGYDFDNYSKMWVRN---OPTINAGAMITY-----LDENGKLIIKASYNSOAEYPFTS	1240
Ddb	I : I :	:
QY	2295 vftntpdkfyfhqhntldenfeigesin----yrgwldlde----krtyftid-eyaiaat	2343
Ddb	I : I :	:
RESULT 14		
AAM68388		
ID	AAM68388 standard; Protein; 2366 AA.	
XX		
AC	AAM68388;	
DT	07-DEC-1998 (first entry)	
DE	Clostridium difficile toxin B.	
EX		
KW	Antitoxin; vaccine; cytotoxin; toxin B: intoxication; immunogen;	
PX	pseudomembranous enterocolitis.	
OS	Clostridium difficile.	
PN	WC9808540-A1.	
PD	05-MAR-1998.	
PF	28-AUG-1997; 97WO-US15394.	
PR	28-AUG-1996; 96US-0704159.	
XX	(OPHI-) OPHIDIAN PHARM INC.	
PA	Thalley BS, Williams JA;	
PI	WPI; 1998-230234/20.	
DR	N-PSDB; AAV30561.	
XX		
PT	Host cell containing recombinant expression vector encoding	
PT	Clostridium botulinum type B or E toxin - useful to treat humans	
PT	and other animals at risk of intoxication with clostridial toxin	
XX		
FS	Example 18; Page 241-249; 428pp; English.	
XX		
CC	This is the amino acid sequence of Clostridium difficile toxin B,	
CC	deduced from the coding region (see AAV30561) of the toxin B gene.	
CC	Fragments of the toxin B gene have been cloned into various	
CC	prokaryotic expression systems, and assessed for the ability to	
CC	express recombinant toxin B protein in E. coli. It would be	
CC	advantageous to use simple and inexpensive prokaryotic expression	
CC	systems to produce and purify high levels of recombinant toxin B	

	for immunisation purposes. The invention specifically relates to
CC	(see patent documents derived from Clostridium botulinum toxins
CC	(see AA68389-400) and their use as immunogens for the production of
CC	vaccines and antitoxins.
XX	
SQ	Sequence 2366 AA;
	Query Match 3.3%; Score 222; DB 19; Length 2366;
	Best Local Similarity 20.4%; Pred. No. 0.00017;
	Matches 220; Conservative 155; Mismatches 401; Indels 304; Gaps
Qy	270 NRAYSDYDASNIENXGLTADT-----WYRKQLIKDCITWTDSK-----ETDMRPL 317
Db	1459 nipyfsvd-segkengflngstkeglfvselpdvvlslskvydmddskpsfygsnnlkdkv 1517
Qy	318 MWMPN-TLTAQYLNMKHGKLPSALPTFFNDAAPAEIN--HYSEIVQQNIERRISE 374
Db	1518 vitkdvniltgyrk-----ddikislsitldektikinswhidesvaeilkfmnr 1571
Qy	375 TGNFDWLRLIHLHDFVYNPNMWNKDSENWFSGIOFGQFLKYENSULTPVANSDYRLGR 434
Db	1572 kgnintsdsim-sfl-----esmniks-----fvnflgsnlkfildanfilsgt 1615
Qy	435 MPINKQDTYRGQEFLLANDIONSPVPVQAOLNWLVLNFGFITANND-----484
Db	1616 tsi-----ggreficendnlqp-----yfkntletcnytlvgnrgqumi 1656
Qy	485 -QANFDSVRVDAPONTADLMNTAQDYFAAAYGMSDASVNKH1---NILEDNNHADPEY 540
Db	1657 vepnyd---lddsdgssstvfhsqyl---yglds---cvakvispnlytdeintpy 1708
Qy	541 FNKIGNPOLTMDDT--TKNSLNHGSLDATNR--WGDLAIHQSLADRENNSTENVVFPNY 596
Db	1709 etantypevivldanyinekinvndlsiryywsndgndfilmastseenkvsgvk1--1765
Qy	597 SFVRAHNNSSODIQNAIRDVTGDKYHTFTPBDEQKGIDATIQDNSTVKKNLYNPAS 656
Db	1766 -----rfnvfkdtianklstnfskq-----dvpvs 1793
Qy	657 YAILLTNKDTIPRYVGd-----LYTD-----GGGYMEHQDPYR-DT 692
Db	1794 eill-----sfpsyyedgligdydlglvalynekyfnngfgmmvsglllyindslyfkpp 1848
Qy	693 LTNLAKSRV-----KY-----VAGGSQMOTNSVGGNNILTSVRYGKGAMTATDGTDET 743
Db	1849 vnmlitgvtvgdkkyfnpingaa---sig---etiiddknnyfqsgvglqgvfste 1902
Qy	744 TGQIGVVVSNT--PNLK-----LGVNDKVVLHMGAAHKNQYRAAV-LTTTDCGVINTS-ID 796
Db	1903 ddfkyfapantildenleagealdftgiidenlyyddnytgavewkeldegmyfsfet 1962
Qy	797 GAPVAMTDENGd---LYLSLHNLYVNGKEEADTAVQGVANPDVS--GYLVAMPVPVGAASNQ 852
Db	1963 gkafiglnagidkyxyfnsdgmkgfvsindnkhyfdsdgymkvgyetei-----dgk 2015
Qy	853 DARTAPSTEKNSGNSAYRTNAAFDSNVIFEAFSNFYPTTKESERANYRIANONADFASL 912
Db	2016 hfyfaengemqig-----vfatedgfykfahhnediqn----eegeeiSYS 2057
Qy	913 GPTESEMAPPQYNSSKDRFLDSTDINGYAFTDRYL-----CMSEPKNYCTDEDLRNAIQ 967
Db	2058 gilnfn-----nklyyfdsstavvgwkledgskyyfidtaeay-2098
Qy	968 ALHKAGLOVMADVDPQYINLPG-KEVATVTRVDDRGNWKDA-----INNLLVYV 1018
Db	2099 ----lglsind--gqyyifnddglmvgvfr-indkvfyfsgdisgesgvnidnfyf 2151
Qy	1019 NTIG---GEYCKYKYGAFDLKQKLYPEITFKKQYSTGVADIPS--OKITEWSAKYPNG 1073
Db	2152 ddnglvqlgvfdtsedykyfapantvndniy-----ggaveysglvrvg-e-dyvyfge 2203
Qy	1074 TWILHRGSGYVLKADGGQYNNIGTTTKQFIPLIQTGKKGGNBGFVKGDNGNYFYDLAG 1133

Db 2204 tytietywiydmenesdkyy-----fnp-----etkackginiiddixyif-dekg 2249

QY 1134 NNAVKTEFTEDSVGNWFFDQCKM-----VENKHFDVDSYGEKGTYYFFLKNGVSPRG 1187

Db 2250 --imrtglisfennnyfngemqfyiniedkmfy----fgedgv-----mqig 2294

QY 1188 LVOTDNGTYTFDNYGKVRN---OTINAGMIYT-----LDENGKLIKASYNSDADYPTST 1240

Db 2295 vftptdgyfahqntldenfegein-----ytwlwlde-----kyyftd-eyiaat 2343

RESULT 15

AAU37017

ID AAU37017 standard; Protein: 5795 AA.

XX AAU37017;

XX AAU37017;

DT 14-FEB-2002 (first entry)

XX Staphylococcus aureus cellular proliferation protein #1187.

DE Antisense; prokaryotic cellular proliferation protein;

XX antibiotic; antibacterial; drug design.

KW Staphylococcus aureus.

OS WO200170955-A2.

XX 27-SEP-2001.

XX 21-MAR-2001; 2001WO-US09180.

XX 21-MAR-2000; 2000US-191078P.

PR 23-MAY-2000; 2000US-206848P.

PR 26-MAY-2000; 2000US-207237P.

PR 23-OCT-2000; 2000US-242578P.

PR 27-NOV-2000; 2000US-253625P.

PR 22-DEC-2000; 2000US-257931P.

PR 16-FEB-2001; 2001US-269308P.

XX (ELIT-) ELITRA PHARM INC.

PA Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GU;

XX Yamamoto RT, Xu HH;

PI WPI; 2001-611495/70.

XX N-PSDB; AAS4876.

DR New polynucleotides for the identification and development of

XX antibiotics, comprise sequences of antisense nucleic acids -

XX Example 3; Seq ID No 12610; 511pp; English.

XX The invention relates to antisense inhibitors of genes essential to

CC prokaryotic cellular proliferation, their use in identifying the

CC genes, their use in the discovery of novel antibiotics, the essential

CC genes themselves and the encoded proteins. The prokaryotes used are

CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella

CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The

CC invention is also useful for the identification of potential new targets

CC for antibiotic development. The antisense nucleic acids can also be used

CC to identify proteins used in proliferation, to express these proteins,

CC and to obtain antibodies capable of binding to the expressed proteins.

CC The proteins can be used to screen compounds in rational drug discovery

CC programmes. The antisense nucleic acid sequence is also useful to screen

CC for homologous nucleic acids which are required for cell proliferation in

CC a wide variety of organisms. The present sequence represents an

CC essential prokaryotic cellular proliferation protein.

CC Note: The sequence data for this patent did not form part

CC of the printed specification, but was obtained in electronic

CC format directly from WIPO at

CC ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 5795 AA;

QY Query Match 3.38; Score 220; DB 22; Length 5795;

Db Best Local Similarity 18.4%; Pred. No. 0.00077;

Matches 281; Conservative 184; Mismatches 534; Indels 532; Gaps 72;

QY 19 SAVSGFDTTKLTNDQYQALNGQLOVLLRFSKAADGNPSGDNTVTQDFS-----KNVATT 73

Db 2277 agiaagtngitvtagtnpad-tiqvv---atggsgetisdegrsddftvvpapnnt 2332

QY 74 ---GGNFD-----YVKGNGQVEFGSWHATNQSNDKDSQMIIVL 109

Db 2333 kiwngghiditpnpsghlinptqamdiaytekmngngaeh--xtinvvggnqngwia- 2389

QY 110 VNGKEVKRQLVNDTKEGAAGFNNDVYKVNPAI-----ENSSMSGFQGLITLPVT--- 159

Db 2390 ---nkpdyytldahgtkvtnantv-kpnsaititpkagtgghsasnpsiltapathv 2444

QY 160 -----VK-----NENVQLVHRPSNDVKTGEG--NYVDFWSELMPVKDSFOKGN 200

Db 2445 ntteivkdygsvntaeainnavgvankrtatikngtamptnlaggstttipvtvyndgs 2504

QY 201 GPLKQFGLQT-----INGQYVYIDPTTGPQRKNFLQSGNNWY----- 239

Db 2505 teevqesiftkdkrelitaknhldpvtgdgkpgtitynnaihnaagqintakteaq 2564

QY 240 -----FSDTGVGTNALQLQFAKGTSSNBOYNGNNAYSYDD----- 277

Db 2565 qvinneratpqqsdahtkvraaqtlineakallnked--nsqlvtsknnlqssvnpv 2622

QY 278 -----KSIENV-----NGYLTADTWYRQPKILKDGTTWDSKET 311

Db 2623 sttgmtqqsldhynakkreaeteitaaqrvindgdata-----qgisidentl---agrn 2673

QY 312 DMRPTLMVWNPNTLQAYYL---NYMKQHGNLLPSALFFPNA-----DAD 353

Db 2674 dl-----tqisqatnlaavesvkqsansldqamgnlqtaindksgtlasqnfldad 2725

QY 354 PAELNIYSEIVOONTEKRISETGNTDMLFTLMHDFVTN-----NPMWN----- 396

Db 2726 edkrnayndav-snaetliinktgtpntaktaveqalnnvnnsakhalngtqnlannakaai 2784

QY 397 -----KQSENVNFGIOFQG-----GFLKYVENSGLT----- 422

Db 2785 talngasdingkqkdaikagangagrvsnagdvqvrnatelntamgtlkhaiaadkintlas 2844

QY 423 -PYANSD-----YLLGRMP-----INIKDQTYRGQEFLLIA 452

Db 2845 skyvnadstkqnayttkvinaehiisgtptvtvtpsevtaaangvnsakqelngderlrv 2904

QY 453 NOIDNSNPVQA-----EQLNWLYLLNFGTITANNDOANFDSVRYDAPNI 499

Db 2905 ak-gnantaiddaltqntpqkakikeqvgganrledigtvqtnngaln----- 2951

QY 500 DADLMNTADYFNAAYGMSDAYSNNKHINILEDWNHADPE-----YFNKITG 545

Db 2952 -----namgildrsianettvkasqnytdasponqstynsavsnaqiingtn 2999

QY 546 NP-----QLTMDPTTIKNSLMHGLSDATNRWGLDAIVHQSLADRENNSTENVIPNYSFVR 600

Db 3000 nptmdtsaitqattgvnnakngingaen-----lrnaqntakqnl-----nti 3042

QY 601 AHDNNSQ-----DQIQNA--IRVDTGKDYHTTFEDEQKIDAYIODQNSTVKYKLYNI 653

Db 3043 shltnnqksaissqidraghvsevtaknaateintqmgnlqefahdqntvqk----- 3095

QY 654 PASVAILLTKKOTIPRVIYGDLYTDGGQYME---HQTTRYDYDTLTNLLKSRKYVAGQSQM 710

Db 3096 ---gvnftdadkkr-----dayfnavsraetlilnktqgantskqdvaaigqvtsakn- 3146

QY 711 QTMVSGVGNNNILTSVRYGKGAWT-----ATDTGTDETRTQGIQVWVSNW----- 754

Db 3147 ---alngqnvtnakntakhalnmltsinnaqkrdltckidqattvavgsntgtqln 3203
QY 755 ---PNLKLGVNDKV-----VLHGAHKNQYRAAVLTITDGVINTSDQGPVAMTDE 805
Db 3204 tamanlqngindkantlasenyhdadsdkktaytqav-tneenltnksngnldkaaven 3262
QY 806 -----NGDLXL-----SSNLVVNKEEDATA-----VQGYANPD--- 835
Db 3263 alsqvtakgalinghnlheqaksnanttinglqhttaqkdklkqvqgaqnavgvtvk 3322
QY 836 -----VSGVLAVVWPVGSNDQDARTAPETKNSGNSAVETNAAFDSNVIFEAFGNF 887
Db 3323 ssantlngamgtlrnsiqdntatnngnyldatesnkn---ynnavdsangvinatsn- 3378
QY 888 VYTPKESERANVRQAQADFFASLGFTSFENAFQYNSKDRTFDSDTNGYAFTRDYD 947
Db 3379 ---pmdanain-----qiatqvtstkna--ldgt----- 3403
QY 948 LCMSEPNKYGTDEDLENAIQA---LHKACLOVMADWVPDQIYNLPGKEVATVTRVDRGN 1004
Db 3404 -----hnltaqkqtatnaidgatnlnka-----qkdalkaqvcs--agrvanvtsiqqtan 3452
QY 1005 VWKDAIINNLVNVNFGGGEYQKYGGAFLDKLQKLYDEIFTKKQV-----STGYA 1056
Db 3453 elnta-mgqlqhgiddenaatkqkyrdaeqkktaydgaavaaakailnkqsgnsdkaa 3511
QY 1057 IDPS-QKITETWSAKYFNGNIIHRGSGYVLKADGGQYVNLGTTTQFLPQIQTGEEKQGN 1115
Db 3512 vdralsqvt-----stkdalngdaklaeakaaakqnlgtln-----hitnaqrta 3557
QY 1116 EGFVKGN---DGNVYFDLAGNMVK---NTF-----IEDSVGNWYFFDDQDGKMKVKNKHF 1163
Db 3558 egqinqattvgd-----vntkvkntautldgamuslqgsin-----dkdatl-rnqny 3603
QY 1164 VDVSDEGKGYFFLKNQYSGFRGLVQTDNG 1194
Db 3604 ldade-skrnay---tqavtaaeaglinlktqg 3630
RESULT 16
ID AAU34143
XX AAU34143 standard; Protein; 2086 AA.
AC AAU34143;
XX AAU34143;
DT 14-FEB-2002 (first entry)
XX Staphylococcus aureus cellular proliferation protein #419.
DE Antisense; prokaryotic cellular proliferation protein;
KW antibiotic; antibacterial; drug design.
XX Staphylococcus aureus.
OS Staphylococcus aureus.
XX WO200170955-A2.
PN 27-SEP-2001.
XX 21-MAR-2001; 2001WO-US09180.
XX 21-MAR-2000; 2000US-191078P.
PR 23-MAY-2000; 2000US-206848P.
PR 26-MAY-2000; 2000US-20727P.
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.
XX (ELIT-) ELITRA PHARM INC.
XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;

XX WPI; 2001-611495/70.
DR N-PSDB; AAS2002.
XX New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX Example 3; Seq ID No 5639; 511pp; English.
PS The invention relates to antisense inhibitors of genes essential to
XX prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 2086 AA;
SQ Query Match 3.3%; Score 219; DB 22; Length 2086;
Best Local Similarity 18.4%; Pred No. 0.00022;
Matches 281; Conservative 183; Mismatches 533; Indels 532; Gaps 72;
QY 21 VSGFDITIKLTNDQYQALNGLOVLLRFSKAADGNPNDVVTDFDS-----KNVATT--- 73
Db 2 iagtngitvtagtfnpad-tiqvv---atqgsgetisdeqrddftvvaqpnhantki 57
QY 74 ---GGNFD-----YKVNQNVQFSGWHATNQSNDKDSQWILVLVN 111
Db 58 wqnghiditpnnpshlntpdamdiaytekmngdaehs--ktinvrvqngnqtia--- 112
QY 112 GREVKRQLVNDTKEGAAGFNNDVYKVPNAI-----ENSSMSGFGQITILPVT----- 159
Db 113 --nkpdvvtldahtgkvtfnantv-kpsaaitipkagthsaessnpsltapathvtnt 169
QY 160 ---VK-----NENVQLVHRSNDVKTGEG---NVVDVFWSELMVPVKDSFQKGNP 202
Db 170 teivkdygsnvtaaeinnavqvantrtatikngtamptnlaggsttltipvtvtyndgste 229
QY 203 LKQFGLQT-----INGQYVYDPTTGQPRKNFLQSGNNWY----- 239
Db 230 evdesiftkadkrelitaknhldpystdgkpkpggttcyqnnalhmaqqintakteaqv 289
QY 240 -----FSDATGVTNALELQFAKGTVSSNQYRNGNAAYSYDD----- 277
Db 290 inneratpqvysdalckvraaqtkineakalgnked--nsqivtsknnlqssvnpst 347
QY 278 -----KSTENV-----NGYLTADTWYRPKQILKDGTTWTDSETDM 313
Db 348 tgmtqgsidvynakkreaeteitaacqvridngdata-----qqisidentt--aqndl 398
QY 314 RPLMWVWPNTLTQAYYL---NYMKQBGHLLPSALPFFNA-----DADPA 355
Db 399 -----tqisqatnlaavesvskgsansldgamnqlqtaindksgtiasqnladaeq 450
QY 356 ELNHYSEIVOQNIETKRISGTGNTDMLRTLHDFVTN-----NPMWN----- 396
Db 451 krnaynqav-snaetlnkqtgtpotaktaveqalnvnnsakhalngtqlnnakqaaita 509
QY 397 -----KDSENVNFGSIQFQG-----GFLYENSIDLt-----P 423

Db 510 ingasdlqgkdaikagangagrsvnaqdvgrnateintamgtlkhaiaadkntlessk 569
QY 424 YANSO-----YRLGRMP-----INIKDQYRGQFLLAND 454
Db 570 yvnadstkqayttkvtaehiisgtpvtvtpsevtaaanqvusakelngderlrvak 629
QY 455 IDNSNFVQA-----EOLNWLXYLLANPGTTITANDQANFDSVRVDPNDIDA 501
Db 630 qnantaialdltqtpqaklikeqvganrllediqvtvgqaln-----674
QY 502 DLWNIAQDYFNAAYGMSDSDAVSNKHNILLEDWNHADPE-----YFNKIGNP 547
Db 675 -----namkgirdsianettvksaqnytdaspanqstynsavsnakgillnqtonp 724
QY 548 -----QLTMDDPIKSLNHLSDAENRWGLDAIVHQSLADRENSTENVIPNYSFVRAH 602
Db 725 tndtsaitqatqvnnaingngaen-----lrnaqtakqnl-----ntlish 767
QY 603 DNNSQ-----DQIQNA--IRDVTGKDYHTFFEDSQKGDIVIQDONSTVKYNYLNIPA 655
Db 768 ltnqksaaisgldraghvsevtaaaknaateintqgnleqahdqn-tvkq-----818
QY 656 SYAILITNKDTIPRYVYGLYDYGQYME---HONRYDITLNLKSRVYVAGQSMQT 712
Db 819 --gvnftdadakr-----daytnavsaetlntkgantksqdvaaiaqvtsakn---869
QY 713 MSVGGNNILTSVRYGKGAMT-----ATDGTDTETRTQIGVYVSNT-----754
Db 870 -alngdqvtnaktakhalnllitsinnakrdlttkiqdattvagveavstgtqlnta 928
QY 755 -PNLKLGVNDKV-----VLHMGAAHKNQYRAAVITVVGVINYSDDGAPVAMTDE--805
Db 929 manlqngindkantlasenyhdaadsdkktaytqatv-naenilnknsngnlkkaavenal 987
QY 806 -----NGDLVL-----SSNLLVNGKEADTA-----VQGYANPD-----835
Db 988 sqvtnakalngnhnleqaksnanntinglqhttagdkllkqvgqagvagnvdykss 1047
QY 836 -----VSGYLAIVVPGASDNODARTAPSTEKNSGNSAYRTNNAFDSNVIFPAFNVFY 889
Db 1048 antlmgamgtlrnsiqdnatnngnqyldatesnkto--ynnavdsaugvinatsn---1101
QY 890 TPYKESRANVRIAQNADFAFLGCFSEFAPQYNSKDRFTFLDSIDNGYATDRYDLG 949
Db 1102 -phmdanaia-----giatgvstlkna--ldgt-----1126
QY 950 MSEPNKYGTDEDLRNAIOA---LHKAGLQVMAWVQDIYNLPKEKAVATVRYDDRGVW 1006
Db 1127 ---hnltaqatnatnaidgatnlka-----qkdalkaqtvs--aqrvanvtsiqgtanel 1177
QY 1007 KDAIINNLLVYVNTIGGGEYQKXYGGAFDLKLOKLYPEITTKQV-----STGVAID 1058
Db 1178 nta-mgqlghgiddenaatkqtkyrdaeqskktaydqaavaakailnqktygsnsdkaavd 1236
QY 1059 PS-QKITEWSAKYFNGNILHRSGYVYLKADGGYVNLGTTTQFLPIOLTGKKGQNEG 1117
Db 1237 raigqvt-----skdalngdakiaecakaakqalglaln-----hitnagrtaleg 1282
QY 1118 FVKGN---DGNVYFYDLAGNVK---NTF-----IBDSVGNVYFFDQDGKVMENKHFVD 1165
Db 1283 qingattvvg-----vntvktntalngamnslggsin-----dkdatl-rnqnyld 1328
QY 1166 VDSYGEKGTFFFLKNGVSFRGLVQTDNG 1194
Db 1329 ade-skrnay---tgavtaaeigilnqktg 1353

RESULT 17
AAU37403
ID AAU37403 standard; Protein: 6281 AA.
XX
AC AAU37403;

XX DT 14-FEB-2002 (first entry)
XX DE Staphylococcus aureus cellular proliferation protein #1573.
XX KW Antisense; prokaryotic cellular proliferation protein;
XX KW antibiotic; antibacterial; drug design.
XX OS Staphylococcus aureus.
XX PN WO200170955-A2.
XX PD 27-SEP-2001.
XX PF 21-MAR-2001; 2001WO-US09180.
XX PR 21-MAR-2000; 2000US-191078P.
PR 23-MAY-2000; 2000US-206848P.
PR 26-MAY-2000; 2000US-207727P.
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.
XX (ELIT-) ELITRA PHARM INC.
XX PA Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX WPI; 2001-611495/70.
DR N-PSDB; AAS55262.
XX New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX Example 3; Seq ID No 12996; 51pp; English.
CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein.
CC Note: the sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 6281 AA;

Query Match 3.2%; Score 212; DB 22; Length 6281;
Best Local Similarity 19.1%; Pred. No. 0.0026;
Matches 249; Conservative 149; Mismatches 505; Indels 402; Gaps 56;

QY 5 ARPDVAAYPSLYNSA--VSGFDTTI-----KLTNDQYQALNGLOLVLLRFSKA 51
Db 2439 aspdkgaytdaynaaknlvngspvntnaadtaatvtaatgngtdnlatakqga 2498
QY 52 ADG-----NPSGDNTVTDQFSKNYATTGGNFYVKNVGNQVEFSGWHATNQ-----97
Db 2499 kdairgmthlsdaqksgitqidsatqvtg--vqsvkdnatnld-----namqnlrnsian 2552
QY 98 -----SNDKDSQWIIIVLVNGKEVKRQLVNDTKEGAAGFNNDVYKVNPAIEN 144


```
Db 2553 kddvkaapdyadrdkqayncavtnaenlnatsptldpsavtqaagvstn----k 2608
QY 145 SMSGFGQIITLPTVYKNEVNLVHRESN-----DYKTEGNYVDWFSELMPVKDSFO 197
Db 2609 talngagml-----ankkqettanlqishlnnaqkqldntqvtnapnl-stvngvkkcae 2663
QY 198 KNGPLKQF--GLQTINGQOYYIDPTTQOPRK-----NFLQSGNNWYFSDTGVGTNAL 251
Db 2664 qldqamerlingidkdkqkqsvftadpektaynnavtaaeail-nqangtnanqs 2721
QY 252 ELQFAKGVSSNEQYRNGNAAYSYDDKSIENW--NGYLTADTWYRPKQILKDCFTWDSK 309
Db 2722 qveaalstvttkqalng-----drkvtaknnaqntlstldlnnaqkgavtgninq 2774
QY 310 ETDMPILMVWPNLTQAYY-----LNYMKQHGNNLPSALPFEN-ADAPAEALNHY 360
Db 2775 ahtaev-----tgaltagelntamgnlknsldkdttgsgqfadadpeknay 2825
QY 361 SEIVOONIEKRISGTGN-----DWLRTLHDFVTNNPWN-----KDSNVNFSIQ 408
Db 2826 neav-hnaenllinkstglnvkdqveaamqvnatakaalngtqnlekakqhantaig1- 2883
QY 409 FQGGFLKYNSDLTPYANSOYKLLGRMPINIKDYRGQEFLLANDINSNPVQA---E 465
Db 2884 -----shlt-----naqkealkqlvqgqsttvaaagqe 2911
QY 466 QLNWLYLLNFETITANNDQANFDSVRVDAPDNIDADLMLNIAQDYFNAAYGMDSDAVSNK 525
Db 2912 q-----kannvdaamdklrqsiaqn-----attkqngy---tdasqnk 2947
QY 526 HINILEDWHA-----DPEYFNKIGNPOLTMDDTIKNSL--NHGLSDAT 567
Db 2948 k-----daynnavttaqgiidqtsptldptvinqaag-----qvstknalngnenleak 2999
QY 568 NRWGLDAIVHQSLA--DRENNSTENWVPIYSEVRHAD-----NNSQOIQNA 613
Db 3000 qg-----asqslgsldlnnaqkqvtvdqngahvdeanqknaqnaqlntamgnlkq 3053
QY 614 IRDVTGKD--YHFFTEFED-----EKGIDAYIDQNSTVKKYNLYNPASYAILLTNKDPI 667
Db 3054 lad---kdatkatvntfdadqakqgqyavntavnaeniiskanggn-----atqaeveqai 3105
QY 668 PRVYVGLYTDGGQYMEHQTRYDYTL---TNLLKSRKVVAGQSQMOTMSVSGNNILTS 724
Db 3106 kqynaakqalngnanvghakdeatalinsndlnqakdalkqvgqnatvavgnnvkqt 3165
QY 725 VRYGKGAWATDTG--TDEFTQIGVVSWTPNKLKGVNDKVVLLHGAABKKNQYRAAYL 783
Db 3166 aqelnnaamtqkkgiaqkeqtkaqgnfnvadpdkqaynqava-----kacellis 3215
QY 784 TTTDGVYNTSDQCAPVAMTDEGDL-----Y 810
Db 3216 atpdrvvtvptaeitaalnkvtqakndlngntlnlatakqnvqhaidqlpnlinqqrdeysk 3275
QY 811 LSHNLV--VNGKEEADTAV-----QYANPDVSGYLVAVPVCASIN-QDARTAPS 859
Db 3276 itqatlvpnvnaiqqaattindamtqkkgiank-----aqikgsenyhdadtqk 3326
QY 860 TEKNSGNS-----AYFTNAFNSNVIFEAFNFFVYPTKESERANVRIAQNADFFPASLG 913
Db 3327 taydnvkaeallkqthntmptntlqqaltk---vndtqalngnqkla----- 3374
QY 914 FTSFEMAPQYNSSKRDFTLSDTDNGYAFTDRYDLGMSPEPNKYGTDEDLRNAIQALHKAG 973
Db 3375 -----dakqaktltlgtldhinda-----dkqa 3397
QY 974 IOVWADNVVDQIYNLPKEVATVTRVDRGNVWVKDAIINNLYVYVNTIGGGEYKKGGA 1033
Db 3398 lttqveqepd-----iatvnnvknqnaqlnnaamtln-----na 3431
QY 1034 FLDKLOKLYPFIYFKKQVSTGVADPSPKITEWSAKYFNGTNILHRGSGYVYLKADGGQY 1093
```

Db 3432 lqdkttlinsinf-----dadq-----akdaytnavshaegilskangsn-- 3473

QY 1094 NLGTTTKQFPIYQLSEKKGQNGEGFKVKGNDGNVYFYFDLACGMVKN 1138

Db 3474 --asqteveqamqrvneakq-----lngndvnrakdaakqvtn 3512

RESULT 18

ABB52677

ID ABB52677 standard; Protein; 1778 AA.

AC ABB52677;

XX 11-FEB-2002 (first entry)

XX Escherichia coli polypeptide SEQ ID NO 749.

XX Escherichia coli; B2/D+A-; antiinflammatory; antibacterial;

KW immunosuppressive; extra-intestinal infection; phylogeny; meningitis;

KW systemic infection; non-diarrhoeal infection; septicaemia;

KW pyelonephritis; antibiotic resistance.

OS Escherichia coli.

XX WO200166572-A2.

XX 13-SEP-2001.

XX 12-MAR-2001; 2001WO-EP03445.

XX 10-MAR-2000; 2000FR-0003145.

XX 02-FEB-2001; 2001FR-0001449.

XX (INRM) INSERM INST NAT SANTE & RECH MEDICALE.

XX Bingen E, Bonacorsi S, Clermont O, Nassif X, Tinsley C;

XX WPI; 2001-550253/61.

XX A library of DNA fragments of Escherichia coli strains for the

PT Phylogenetic determination of a given strain comprises polynucleotides of

PT nature B2/D+ A- -

XX Example 6; Fig 6; 646pp; English.

PS The invention relates to a library of DNA fragments of Escherichia coli

CC strains comprising polynucleotides (ABA88577-ABA88729 and ABA89533)

CC and encoded proteins (ABB52459-ABB52919 and ABB52954-ABB53094) of nature

CC B2/D+A-. The polynucleotides have potential antiinflammatory,

CC antibacterial and immunosuppressive activity as part of pharmaceutical

CC compositions used to treat, palliate or prevent extra-intestinal E. coli

CC infections. The polypeptides are useful for determining the phylogenic

CC group of a given E. coli strain. These polypeptides can detect and treat

CC an undesired development of E. coli, particularly an extra-intestinal

CC infection that include systemic and non-diarrhoeal infections such as

CC septicaemia, pyelonephritis and meningitis this is particularly

CC advantageous as bacterial resistance is increasing with the more

CC frequent use of broad spectrum antibiotics.

XX SQ

Sequence 1778 AA;

Query Match 3.2%; Score 211; DB 22; Length 1778;

Best Local Similarity 19.4%; Pred. No. 0.00053;

Matches 262; Conservative 171; Mismatches 519; Indels 402; Gaps 64;

QY 15 SLXNSAVSGEDTTIKLTN-----DOYQALNGQLQVLLRFSKAADGN---PSGDNVYDQF 66

Db 261 alnaqasgesialqntseaeqnaialqggsiaakvnsialgsnsssgenaia--l 318

QY 67 SKNVATTGGNF---DYVKYNG-----NQYFSGWHATNOSNCKDSWII 107

Db 319 gegsaaggsnslafgsgsranngndsvaigvgaataatnsvaigagsttdasn-----t 371

Db 1307 neirnlksysfsgaggtysll-----ssypstlnlnlskdldwlnfnidnevreis 1358
QY 110 VNGKEVKQLVNDTKEGAAGFNRRNDVYKVNPAIENSSMGFGQ-----IITLPVTV 160
Db 1359 eugtakkklkv-----lskdlnkllgnqtid-fsgdlnkdryiftceldd 1411
QY 161 KNE---NVOLVHRFENDVKTEGNTVDFWSELMPVKDFQKNGPLKGFGLQTINGQOYY 217
Db 1412 kisllelnlvaksyllsgdknyl-----lsnlsntiek-----intlgidsknlayny 1462
QY 218 IDPTGQPKNFLLQSGNNWYFDSDTGVGNALQLQAKGTVSSNEQVRCNGNAAYSD 277
Db 1463 td-----esnnk--yfgaisktsqksi-ihykkdsknillefyndstlefnskd 1507
QY 278 KSIENWNGYLTADTWYRPKQILKOGTTWTDKSETDWRPILWVWNPNTLTQAYYL----- 331
Db 1508 faedinvmkddi-----ntitgkyvvdnntdk 1536
QY 332 -----NYMKOHG-----NLLPSALPFF-NADADPAELNHY5----- 361
Db 1537 sidfsislvsknqkvnglylnesvyssylfvdksnsgdhntsnfmnlfdnlsfwklfg 1596
QY 362 -ELVQONIEKRISSETGNT-----DMLRTLMHDFVTN-----N 392
Db 1597 feninfvdkyflvgktulgyveficdnknkaidyfgewktsesskstifsgngrvvve 1656
QY 393 PMWKKD-----SENWFS-----GIOFOGFLKYENSDL-----TPYANSDY 429
Db 1657 piypntgedistsldfsyeplygid-----rylnkvliapdlytslnintnyysney 1710
QY 430 -----RLGRMPINIKQTY-----BGOEFLANDIDNSN-PVVOAEQLNWLYY 472
Db 1711 ypeiivlnpntfhkvnlnldsssfeykwstegsdilvryleesnkklqkirkigi-- 1768
QY 473 LLNFGTITANNQANFDSVRVDPADPNIDDLNIAQDY--FNAAAYGMSD-----AVSNK 525
Db 1769 -----lsntq-sfnksnidfkdkkkslylmsnfksfnseiseld-dhlgfkliidnk 1819
QY 526 H-----INILEDWNHADPEYFNKIGNPQLTWDDTTKNSLNHGL----- 563
Db 1820 tyvydesklvkglinlnnsalfydgiefnlvgtwtingkkyfyfintgaaltsykiin 1879
QY 564 -----SDATNWGL-----DAVHOSLADRENNSTENVPIPNYSFVRAHDNSQOIQN 612
Db 1880 gkhfyfndgvmglvfkpgdgfeyfapantqnnilegqal-----vyqg 1924
QY 613 AIRDVRGKYHTTFDEQKGI-----DAVIOQNSTVKKYNIYNTIPASVAILLTK 664
Db 1925 kfitlmgkky--fdnnskavtgrwlinnekyfnpnaiaavgl-----qvldnnk 1974
QY 665 DTIPRVYGP-----LYTDGQYMEKOTRYDITLNLKSRVKYVAGQSQMOTMSVGGNNN 720
Db 1975 -----yyfnpdtaiskgwtvngsryfddtaafngyk-----tidgkhf 2017
QY 721 ILTSVRVYKGMATADTG-----TDETRTGIGVWVS-----NTPNLKLGVDNDKV 766
Db 2018 yfddcwvklgfstngsfeyfapantynniiegqaiyvyskfltingkkyfyfndnaskav 2077
QY 767 LHMGAAHKNQOY-----RAAVLTITTDGVINTYSDGAPVA---MTDENGDLVLSHSHN- 815
Db 2078 tglqtdskkyfyfnttaeaatgwtidgkkyfyfnttaeaatgwtidgkkyfyfntta 2137
QY 816 -----LVVNKEEADRAVOQYANPDVSVGLVWVPVPGASDNQDARAPSTKNSGNSAY 869
Db 2138 iastgytlingk-----hryfntd--gimqigvfkpg--pugfeyfapa---ntdanni 2193
QY 870 RTNAFDSNVIFEAFSNEFYPTKESERANVRIAQNADFF-----ASLGFTSFEMAPQ 922
Db 2194 egqallynefitingkkyfyfsgdsakvtgrwlinnkkyfnpnaiaaahltci----- 2238
QY 923 YNSKDRFTLSDINDGVAFTDRYDLGMSPEKNKYGTDEDLNATOALHKA--GLQVMAOW 980
Db 2239 -nddkyysydgllqngyftlier-----unfyfdanneskmvtgfvkpgngfeyfa-- 2288

RESULT 20

AAW68387
ID AAW68387 standard; Protein; 2710 AA.
XX
AC AAW68387;
DT 07-DEC-1998 (first entry)
XX
DE Clostridium difficile toxin A.
XX
KW Antitoxin; vaccine; cytotoxin; toxin A; intoxication; immunogen;
KW pseudomembranous enterocolitis.
XX
OS Clostridium difficile.
XX
PN WO9808540-AL.
PD 05-MAR-1998.
XX
PF 28-AUG-1997; 97WO-US15394.
XX
PR 28-AUG-1996; 96US-0704159.
XX
PA (OPHI-) OPHIDIAN PHARM INC.
XX
PI Thalley BS, Williams JA;
XX
DR WPI; 1998-230234/20.
XX
DR N-PSDB; AAW30560.
XX
PS Host cell containing recombinant expression vector encoding
PT Clostridium botulinum type B or E toxin - useful to treat humans
PT and other animals at risk of intoxication with clostridial toxin
XX
XX Example 15; Page 220-230; 428pp; English.
XX
CC This is the amino acid sequence of Clostridium difficile toxin A,
CC deduced from the coding region (see AAW30560) of the toxin A gene.
CC Toxin A is a potent cytotoxin that plays a direct role in damaging
CC gastrointestinal tissues. Severe cases of C. difficile
CC intoxication result in pseudomembranous colitis. This would be
CC prevented by neutralising the effects of toxin A in the
CC gastrointestinal tract. Examples are provided of the production
CC of recombinant C. difficile toxin A in host cells and of the in
CC vivo neutralisation of toxin A by antibodies against recombinant
CC toxin A polypeptides. The invention specifically relates to
CC recombinant proteins derived from Clostridium botulinum toxins

CC (see AAW68389-400) and their use as immunogens for the production of
CC vaccines and antitoxins.

XX Query Match 3.1%; Score 209; DB 19; Length 2710;
XX Best Local Similarity 18.1%; Pred. No. 0.0013;
SQ Matches 263; Conservative 198; Mismatches 512; Indels 484; Gaps 73;

QY 60 NVTVQFSEKNTATCGNEDYKVGNGVFEFSWR--ATNQSNDKDSOWI-----ITLPTVV 160
DB 1307 neirnklsydgaggtysll-----lsyptstlniskdwlwifndnevreisi 1358
QY 110 VNGKEVKROLVNDTEGAAGFNRNDVYKVNPAIENSMSGFQ-----ITLPTVV 160
DB 1359 engtlkkgklidv-----lskldinknllignqtid-fsgdldnkdryifitceldd 1411
QY 161 KNE---NVQLVHRFNDVKTGEGNVDFWSELMFVKDSFOKNGPLKOFGLQTNGQYY 217
DB 1412 kislilelnlvaksyllsgdknlyl-----fslnsntiek-----intlgidsknlainy 1462
QY 218 IDPTGQPKNELLOSNNWIFVSDTGTGVTNLELOFAKGTVSSNEOYRNGNAAYSD 277
DB 1463 td-----esnk--yfgalsktsqksi-ihykkdsknilefynostlefnskd 1507
QY 278 KSIENVNGYLTADTWYRPAQILKDGTTWTSKETDMPILMWMPTLTQAYYL----- 331
DB 1508 fiaedinvfmkddi-----ntitgkyvdmntdk 1536
QY 332 -----NYMKQHG-----NLLPSALPFF-NADADPAELNHS----- 361
DB 1537 sidfsislvsksqkvnglylnessyylfdvksdghntsnfmnifldnlsfwklfg 1596
QY 362 -ETVQONIEKRISGNT-----DMLRTLMHDFVN-----N 392
DB 1597 feninfvdkyftlvqktnlygvefidcnknidlyfgewktskskfsgngnrnvve 1656
QY 393 PMWNKD-----SENWFS-----GIOFGGFLKYENSGL-----TPYANSXY 429
DB 1657 piytpdtdedistsldfsyelygid-----rylnkvllapdltytslntnyysney 1710
QY 430 -----RLGRMPINKOQTY-----RGOELLANDDNSN-PVQAEQLNWLXY 472
DB 1711 ypeilvlnputfhkknvlnldsssfeykstwsgdfilvryleesnkklqkirkgi-- 1768
QY 473 LLNFGTTTANNDQANFSDVRDAPNDIDADLMIAQDY--FNAAYGMDS-----AVSNK 525
DB 1769 -----Isntq-sfkmsidfkdkkkslsgyimsnfksfnseneldrdhlgfkliidok 1819
QY 526 H-----INLEPNHADPEYFNKIGNPOLTMDDTIKSLNHEL----- 563
DB 1820 tytydedsklvkglinnalfyfdpiefnlvtgwtlqngkkyfyfintgaaltsykiin 1879
QY 564 -----SDATNREL-----DAIVHOSLADRENSTENVIPNYSFVRAHDNNSQDQION 612
DB 1880 gkhfyfnddgvmqlgvfkpgdgyfeyapantqnnilegael-----vyqgs 1924
QY 613 AIRDVTGKYHTTEFEQRGI-----DAYIQDNSTVKKYNLYNPASVAILTNK 664
DB 1925 kftlqngkky--fdnkskavtgrwriinnekyfnpnnaaavgl-----qvidnkn 1974
QY 665 DTIPRVYGD-----LYTDGGQYMEHOPRYDVTNLKSRVKYVAGQSQMTMSVGGNNW 720
DB 1975 -----yyfpdpdtaiskgwtngsryyfdtdlaiafngyk-----tidgkhf 2017
QY 721 ILTSVRVKGKAWFATDYG-----TDEFTQIGVVVS-----NTPNLKLGVDNDKV 766
DB 2018 yfddcvvkiqvfstsgfeyfapantynnnilegavyskfltlngkkyfyfndnkskav 2077
QY 767 LHMGAHKNQOY-----RAVLTTTDDGVINTSDGAPVA---MTDEGNDLYLSSHN- 815
DB 2078 tgiqtidaskkyfntntaeaatgwtidgkkyfntntaeaatgwtidgkkyfntnta 2137

QY 816 -----LVVNGKEADTAVOGYANPOVSGYLAWVPVVGASDQADARTAPSTEKNSNSAY 869
DB 2138 iastgytllngk-----hfyntd--gimqigvfkp--pngfeyfapa---ntdannl 2183
QY 870 RTNAPDSNVIFEAFSNFVYTPTKESBRANVRVIAQNADEF-----ASLGFTSFEMAPQ 922
DB 2184 egqaillynefltingkkyfygsdskavtgrwriinnkkyfnpnnaaiahletl----- 2238
QY 923 YNSKDRFTFLDSTINDNYATDRYDLGMSBPNKYGTDEDLRNAIQALHKA--GLQVMADW 980
DB 2239 -nndkyyfsydgilqngyitier-----nnfydanneskmvtgfvkpgpngfeyfa-- 2288
QY 981 VPDOYINLPKGEVATVTRVDDRGVNWKDAILNNLXYVNTIGGSEYOKKY----- 1030
DB 2289 -panth-----nnieqgalvyqnkfil--ting-----kkyfndskavt 2326
QY 1031 GGAFDLKLQKLYPBIFTKKQVSTCAID-----PSQKITWSA-----KYFNGTNIL 1077
DB 2327 gwtidg-kkyfnnlntaeaatgwtidgkkyfnnlntaeaatgwtidgkkyfntntf 2385
QY 1078 HRGSGYVLKAGGOYNNLGTGTFQLPQLTGEKKQNGEFGVKGNDGNYF-----YD 1130
DB 2386 lastgwt-singkhfy--fntdgmgi-----gvfkpgpngfeyfapantdann 2430
QY 1131 LAGNMV--KNTFIEDSVGNWYFFDQDGKMWENKHEFVDVDSYGEKGTFFLKNGYSEFGL 1188
DB 2431 iegqaillyqnkfiln-gkkyfyfsgdsdskavtglrtidgkky-----yfnntavavlgw- 2483
QY 1189 VQTDNGTYTFDNYCMKWRN--QTINAGAMIYTLDBNGKLKIKASYNDSA----- 1234
DB 2484 -qtingkkyfntntasiasgtlylsgkhfy-----fntdgmgiqgvfkpgpdg 2530
QY 1235 -BY--PTSTDVGKMLDQ 1248
DB 2531 feyfapantdannieqg 2547
RESULT 21
AAR99393
ID AAR99393 standard; Protein; 2353 AA.
XX
AC AAR99393;
XX
DT 15-JAN-1997 (first entry)
XX
DE Haemophilus adhesion protein HA2.
XX
KW Haemophilus adhesion protein; HA2; hsf protein; vaccine.
XX
OS Haemophilus Influenzae type b strain C54.
XX
PN W09630519-A1.
XX
PD 03-OCT-1996.
XX
PF 22-MAR-1996; 96WO-US04031.
XX
PR 24-MAR-1995; 95US-0409995.
XX
PA {UYSL-} UNIV ST LOUIS.
XX
PA (UNIW) UNIV WASHINGTON.
XX
PI Barenkamp SJ, St Geme JW;
XX
DR WPI; 1996-455364/45.
XX
DR N-PSDB; AAT41476.
XX
PT Recombinant Haemophilus adhesion proteins HA1 and HA2 - for use in
XX
PS vaccines against H. influenzae infection.
XX
PS Claim 5; Page 66-73; 120pp; English.
XX

CC Haemophilus adhesion protein HA2 (AAR9393) is associated with the
 CC formation of surface fibrils involved in adhesion to various host
 CC cells; it is also referred to hsf (Haemophilus surface fibrils).
 CC Its amino acid sequence was deduced from a genomic DNA clone
 CC (AAT41476) derived from Haemophilus influenzae type b strain C65.
 CC Large quantities of recombinant HA2 can be produced in transformed
 CC prokaryotic or eukaryotic host cells, for use in vaccines against
 CC H. influenzae infection.

XX
 SQ Sequence 2353 AA;

Query Match 3.18; Score 208; DB 17; Length 2353;
 Best Local Similarity 19.78; Pred. No. 0.0012;
 Matches 282; Conservative 156; Mismatches 447; Indels 548; Gaps 74;

QY 29 KLTNDQ--YQALNGQLOVLLRFSRAADNPSSG-----NTVTDFSKNYATTCGNFD 78
 DB 1088 klvnaeglataln-nlswtakadyadgesetdqvkgakvtfkagkn----- 1137
 QY 79 YKVVNGNQVEF-----SGWHATQSNQDKDSQWLIIVLVNGKEVKRQ 118
 DB 1138 -lkvkqsekdfyslqdtltgtsitlsgtangrndgtgvinckdg-ltitlangsaagt 1195
 QY 119 LVND-----TKEGRAAGFNRN-----DVKY-----VNPATENSSMSGFGQ 152
 DB 1196 asngntisvkgdigaagknkeltvnsalktykdtantadtdqkdfhaavknanevefv 1255
 QY 153 IITLPVTKRNVNQLVHRFSNDV---KTGEGNYDFWSELMPVKDSFGKNGPLKQFGLQ 209
 DB 1256 kngatvsaktin-ngkhtvtldvaeekvgdglekd-----tdgkiklkvdn 1300
 QY 210 TINGQYYIDPTGQPRKFLQSGNNWIFDSDGVTGNALQLQFAKGTSSNEQYRNG 269
 DB 1301 tdgnnlltvdatkgas-----vakgefnavttdattaqgnaue-----rgkvvvkgs-ng 1350
 QY 270 NAAYSYDSSIEVNGYLTADWIRPKQILKDGTTWDSKETDMRPILMVPWPNLTQAY 329
 DB 1351 atatedkktvatvgvaka-----indaatfvkveddsati----- 1388
 QY 330 YLYNMKQHGNIPLSALPFFNADAPALNHYSEIVQONIEKRISETGNTDWTLMHDFV 389
 DB 1389 -----ddsp-----tddgandalka--gdtl 1407
 QY 390 T-----NPPANKDSENVNPSGTFQCGFLKYENSDLTPYANSYRLLGRWPNIKDTYR 445
 DB 1408 tikagknlkvrdgknitfa----- 1427
 QY 446 GQETLLANDIDNSNPFVQARQLMNYLLNFGTITANNQANFDSVRVDAPDNIDADLAN 505
 DB 1428 -----landisvksatvsdk-----lsigt--ngknvnits-----dtkgin 1462
 QY 506 IAQDYNAAYGHDSADVSNKRIINLEDWNAHDPYENKIGNPOLTMODYTKNS-----L 559
 DB 1463 fakd--sktgdd-----anihngias-----titdtilnsgattnl 1498
 QY 560 NHGLSDATNR-----NGLDAIVHQSLADRENNSTENWVTPNVSFVRAHNNNSQ 607
 DB 1499 ngitdhekkraasvkdvlagnvrvkpas-----annqveni-----dfvatyd---- 1545
 QY 608 DQIGNAIRDVTG-KDYHITFE--DEQKGDIDAYIQDNSTVKKYNLYNPASVAILTNK 664
 DB 1546 -----tvdfvsgdkdtsvteskdngkrtvekvikgatsvikdhn-----gk1ftgk 1592
 QY 665 DTIPRVYIGDLYTGGQYMEHQRYRYDTLNLKSRKYVA---GGQSMOTMSVGGNNNI 721
 DB 1593 e-----lkdanngvgtvte--tdgkdegnglvtakavidavnkagrvkttgangannd 1644
 QY 722 LTVRYKNGKATNDFTG-----DETFTQIGVVVVSNTPNKLGVNDKVLVLMGAH 773
 DB 1645 fatvasgntvtfadngttaevtkandgsvtkvynkvad--glkl-dgkivad----- 1696
 QY 774 KNOQYRAAVLTDTGVTINYTSDQAPVAMTMDENGDLVLSHNLVYVNGKEADTAVQGYAN 833

DB 1697 -----ttvltvaagkvt-----ap-----nngd-----gkfkvdas----- 1722
 QY 834 PDVSGYLAVVVPVGCASD--NQDARTAPSTKNSG-----NSAYRTNAFDSNVLFBAFSN 886
 DB 1723 -----gladalnklswtatagkegtgevdpausagdevkagd-kvtfkagd- 1767
 QY 887 FVYPTTKESERANVRIAQNA--DFEASL-----GFTSEMAFOYNSSMDRTFLDST--IDN 938
 DB 1768 -----nlkiksgkdfctyslkkelditsve-----fkdganggtgsestktkd 1811
 QY 939 GYAFTDRYDLGMSNPKNYGTDEDLRNLAIQALHKA-----GLQVMADWVQQIYNLPGE 992
 DB 1812 gltitpangagaagaaantantistvtkdglisagknkavtnvsglkkfgdg-----htlangt 1866
 QY 993 VA-----TVTRVDDRGVWKAIDALINNNLYVNTTGG-----GEYOKKY 1030
 DB 1867 vadfekhydnaykdltnidekgadnnpvtadntaatvdydlrglqvwisadktgpepuey 1926
 QY 1031 GGAFLDKLQKLYPEITFTKKQVSTGVAIDPSOKITEWSAKYFNGTNIL--HRSGGYVLK-- 1086
 DB 1927 naqvrnane-----vfksgnginv-----sgktingtrvitfelakgevvksn 1970
 QY 1087 -----ADGGQ-----YYN-----LGTTRKQFLP-----IQLTGCKQ-- 1113
 DB 1971 eftvknadsetnlvkvdmvyskediqpatskpmgtkkykvengkvvaangsktevt 2030
 QY 1114 ---GNEGFPVKGNDGNYFYFDLAGNMVKNTF---TEDSVGNVWFFDQDGKVE----- 1159
 DB 2031 ltnksgvytgn-----qvadalaksgfeglgadaaeakafaesakdkqlakdaet 2083
 QY 1160 -NKH-----FVQVDYSGEKGTYFFLKNQVSYF-----RGGDLWQ 1190
 DB 2084 vnahdkvrfanglntkvsaatvestdangdkvtftvtdvelpitqiyntdangnkivk 2143
 QY 1191 TNGYGYFDNKGKVPNTINAGAMIYLDENG-KLIKASVN-SDAEYPTSTD 1241
 DB 2144 kadgkvwelnadgtaskevltig-----nydangkkvkvvtengadkwytytnad 2192

RESULT 22
 AAB01830
 ID AAB01830 standard; Protein; 1222 AA.
 XX AAB01830;
 XX AC AAB01830;
 XX DF 11-SEP-2000 (first entry)
 XX DE H. influenzae strain K1 mature full-length HMW1A protein, SEQ ID NO:37.
 KW Mature HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
 KW non-typeable Haemophilus influenzae; NTH1; non-encapsulated;
 KW recombinant production; Escherichia coli; antibacterial; vaccine;
 KW human disease; Otitis media; epiglottitis; pneumonia; tracheobronchitis;
 KW detection; diagnosis.
 XX OS Haemophilus influenzae strain K1.
 XX FH Key Location/Qualifiers
 XX FT Misc-difference 307 /note- "Encoded by GG"
 XX PN W0200020609-A2.
 XX PD 13-APR-2000.
 XX PF 07-OCT-1999; 99WO-CA00938.
 XX PR 07-OCT-1998; 98US-0167568.
 XX PA 08-DEC-1998; 98US-0206942.
 XX PA (CONN-) CONNAUGHT LAB LTD.
 XX XX

Loosmore SM, Yang Y, Klein MH;

1111

WPI: 2000-303789/

N-PSDB; AAA52180.

Nucleic acid molecule for producing recombinant high molecular weight proteins of Haemophilus which are used as a vaccine to provide protection against Haemophilus induced diseases in humans -

protection against haemorrhagic disease

Claim 8: Fig 20A-R: 307pp; English.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

The invention relates to the recombinant production of Haemophilus influenzae high molecular weight (HMW) proteins in Escherichia coli. The expression construct used to effect recombinant expression comprises a promoter functional in E. coli (e.g., the T7 promoter) operably linked to a modified hmWABC operon from a non-typeable (non-encapsulated) H. influenzae (NTHI). Most HMW-expressing NTHI strains contain two hmw gene clusters termed hmw1ABC and hmw2ABC. Each hmWABC operon comprises hmwA, hmwB and hmwC genes. The hmwA genes encode the structural HMWA proteins and the hmwB and hmwC genes encode accessory proteins which are responsible for post-translational processing and secretion of the HMWA proteins. The modified hmWABC operon used in the expression construct only of the invention contains an A gene modified such that it encodes only the mature HMWA. The invention also discloses hmwa genes (AAAS2175-A52198) and HMWA proteins (AAB01824-B01849) from the non-typeable H. influenzae strains Joyce, KL, K21, LCDC2, PMH1, 15 and 12. The nucleic acids and vectors are used for the production of recombinant H. influenzae HMW proteins which can be used as vaccines to mediate a humoral or cell-mediated immune response to provide protection against diseases in humans caused by H. influenzae (e.g., otitis media, epiglottitis, pneumonia and tracheobronchitis). The HMW proteins are also useful as antigens in immunoassays for detecting antibodies against Haemophilus, HMW proteins and/or HMW peptides. The nucleotide sequences encoding the HMW proteins can be used to isolate and clone hmw genes from other non-typeable strains of Haemophilus via hybridisation reactions. The present sequence represents a mature HMWA protein from a non-typeable strain of H. influenzae.

Sequence 1222 AA;

Query Match	3.1%;	Score 206;	DB 21;	Length 1222;
Best Local Similarity	19.6%;	pred. No. 0.00064;		
Matches 229;	Conservative 147;	Mismatches 460;	Indels 332;	Gaps 54;

QY 73 TGG--NFDYKUNGQVEF-----SGHATNQNDKDSQHIIV 108
||| . | | :
||| . | | :

```
Db 170 tggglsfkylctgnrdnsfethfrgrlnisgkvdilmqargew--nrthwgrshwvnt 226
```

QY 109 LVNGKEVKRQLVNDTKECAAGFNRNDVKVNP-AIENSSMSGFGQIITLPVTYKKNENVL 167

```

Db      227 rlnvsensyfnvltldsgsas-----spgagpnaqsglngl----- 263

```

168 VHRFSNDV--ATGEGNYDVFWSLMPVKDSFQKGNGLKQFGLQTING--QQWYIDPTG 223

```
Db 264 --sfndtvtfniaassavnf-nlppldvktngnhtlfgknlsrlg99gdvfnhinasss 320
```

[illegible]

DB	321	nyqtygvliesqni-sasaggsslkkksegsthaaftlxndllmatggulsinqvagids	377
QW	325	-----AVGQCNAAVGVDDQVSTENYNGVITATGVVPRKAVIYKQGTWTFWDSKETD	332

QY	-----QYRNGNAAYSIDKSTSENVGILITADTWIKPRKLRGGLIWDSABED	31
265	-----	
380	nklttiantknitfeccgnitlaadttoje-kanit-----vkegan-----	420
nh	-----	

DD	380	PIKKSIIANUKHILLESUUNILIAAAMAPRE- LAGNIIIC	-----	vkregum	412
Qd	313	M8PILMUMEPNTIIOAYVINUMKONGHILPSALPEENADAPPAELNHVSEIIVONIEKRI	372		

QY	ANFAYVNNFNINMILNLLNLNNAJONLLEJOFETIYNALDITENEMACATKXAVAGW
CYC	II III ; II III ; III :
pb	-----vtirsanvand-----ksalsirgvntkgnlvtqsaif-niekn! 460

QV 373 SETGNTDWLRTLMDHFEVTNNPMWNKDSENVNFSGIQFGG--FLXVENS---DLTPYANS 427

Dbb 461 tvegsakflapnysfvsvglfdngsknisia---kqgaifkdientgslnitrrksds 516

100

•

QY	428	DYRLGRPTINIKQ--TYRGEFLLANDIDNS-----NPVVQAEQLNWLYLLNNGTITA	481
Db	517	nbht1-----ikgnitnrkgdlntngndteiqgnisqke-----gnli	559
QY	482	NNDQANFD-----SVRVDPADIDALMIADQYENAAIGMDSDAVSKHNILEDWN	534
Db	560	ssdkvniteritikagvngdsngseatsanl-----tikkeiklntdin	605
QY	535	HADPEYFNKI-----GNPQLTMD-----TIKSLNHGILSDATNRWGLDAIVH	577
Db	606	isg---finkaeitakdsnultigdsdagntdakvtfsnvkdsiskasdhvntlnskve	662
QY	578	-----QSLADRENNT-----ENWVINYSEVRAH-----DNNSDQIQNAIRDVTG	619
Db	663	tsgdtdstcdggnungltitaknvtvnnn---ieshktvnltaesenvtkagttnatig	720
QY	620	KOYHTFFDEQKGDIDAYOQNSYVK--KYNLNYI-----PASYALLFNKDTIPR	669
Db	721	svevtaktgdkigiesngvnnitcasgdtlnvsnitqnvvtvaasgavtttkgslna	780
QY	670	VYGGDYITGGQYMEHOFXYDYTLNLKSRVYVAGOSMOTMVSIGVGNNLTSVRYCK	729
Db	781	-----tsgnanittkgeingevksagvnnitcasgdtlnvsnitqnvvtva---ns	830
QY	730	GAMTADTGTDETRTQIGGVVVSNTPNLKLGVNDKVVYLHMGAAHKNOQYRAA	789
Db	831	gaittte-gstlnattgdanittqgn---ingkvessg-----svtliatgqt	876
QY	790	INTSDQAPVAMTDENGDIYLSNNLVNGKEADTAQYANPDVSGYLA-VWVPVGA	848
Db	877	lavgnisgdtvtitackgklittqts-kingtkavttssq---sgdisgtisgntvsaa	932
QY	849	SDMQDARTAPSTKNSGNSAYRTNAAFDSNVIFEAFSNVFTPTKESERANVRIAQNA	908
Db	933	tsalltqagskieakgea---nvtsatgltiggtisgntvntantd---nltkdgari	986
QY	909	FASLGFTSFEMAPQYNSKDRFILTSTIDNGVAFTRDVLGMSGP-----NKYGT--	958
Db	987	katgg---avltatggtlttetsditsnggtlttakdssiagisnaavnlntgtlt	1044
QY	959	-----DEDLRNAIQALHRAQLQVWADWVPDQIYNLPQKEVATVTR	999
Db	1045	tvagskieaasgtlvlnakdaqlgdaasgdhvtvnatna-----ngsgsviatls--	1094
QY	1000	DDRGVNWKDAIINNLYVYVNTIG-----GGEYQKKY-----GGAFLEKLQ	1039
Db	1095	-srvnielgdlitngliniiskntvllkgveidvkiqpgiasvneieak-alekvk	1153
QY	1040	KLYP---EITFKQVSGVAIDPSQKIT	1064
Db	1154	disdeeretlakgsvavrfaepnnait	1181
RESULT 23			
AAB01828			
XX	ID	AAB01828 standard; Protein; 1228 AA.	
XX	AC	AAB01828;	
XX	DT	11-SEP-2000 (first entry)	
XX	DE	Haemophilus influenzae strain K1 full-length HMW1A protein, SEQ ID NO:34.	
KW	KW	HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;	
KW	KW	non-typeable Haemophilus influenzae; NTHI; non-encapsulated;	
KW	KW	recombinant production; Escherichia coli; antibacterial; vaccine;	
KW	KW	human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;	
KW	KW	detection; diagnosis.	
XX	OS	Haemophilus influenzae strain K1.	
XX	FH	Key	
Ft	Ft	Location/Qualifiers	
Ft	Ft	Misc-difference 313	

[illegible]

Qy	265	-----QYRNGRAAYSDDKS	TENWNGVILTADTWPRPQILKDCGTWTTDSKETYD	312
Db	386	nlkksllanknitfeggnitlaadkhpie-ikgnit	-----vkegan	426
Qy	313	MREILMVWPNLTQAYLYNVMKQHGILLP	GAALPFENADADPAELNHYSEIVQONTBKRI	372
Db	427	-----vltirsanynd	-----ksalsirgnvtnkgnltvtgsai--nieknl	466
Qy	373	SETGNTDWLRLTHHDFVTNPNMWNKDS	ENPFGIQGG--FLUKYENS--DETYPYAS	427
Db	467	tvegsakflanpnysfnvsglfdngkgnisia	---kgsaifkdientgslntittksds	522
Qy	428	DYBELGRMPINIKQ--TYRQBELANDIDNS	-----NPVQVQSLWLYLLNFGTITA	481
Db	523	nhhti-----ikgnitorkgdlntnngnd	teidqgnisqke-----gnitl	565
Qy	482	NNDQANFD-----SVRYDAPDNIDADL	MNTAODYFNAYAGMDSDAVSNKHINILEDNW	534
Db	566	ssdkvniteritikagvngdsdsneatsanl	-----tiktkelkitndln	611
Qy	535	HADPEYFNKI-----GNPOLTMD	-----TIKNSLNHGLSDATNRWGLDAIVH	577
Db	612	isg--fokaeitakdnslitgdnadagnat	dakvtfnsvndkslissdhovtlnskve	668
Qy	578	-----QSLADRENST-----ENVVIPNYS	FPRAH-----DNNSODQOINAIROVYG	619
Db	669	tsgdtstedggnantglitaknkvnnon	---itshktnitcasenvttkagttinatgt	726
Qy	620	KDXYHTFFPEBQKIDAYIQDQNSTYK	--KYNLXNI-----PASVAILTNKDTIPR	669
Db	727	svevtaktgdikggleensgnvnit	asgdtlnvsnltgnavtvaasgavttkgtstina	786
Qy	670	VYGDLTGDCQYMEHOTRYDTLTLLKSRVYK	VAGGOSQMTSMVGGNNNLTISVRYGK	729
Db	787	-----ttgmanittktgeinvksasgnvnit	asgnltlnvsnltgnavtva--ns	836
Qy	730	GAMPATGTGDETFQTGIGVVVGNTPNLK	LGVNDKVKVLHMGAAHKNQOYRAALTTTGGV	789
Db	837	gaiftte--gstlnattgdanittqtn	---ingkvessag-----svtliatqgt	882
Qy	790	INVTSDQAPVAMTDENDILYLSHNLVNG	KEADTAVQGYANPDVSGYLA--VWVPVGA	848
Db	883	lavgnisgdtvticadkglkttquss--k	ingtksvtssq-----sgdisgtsignvtvsya	938
Qy	849	SDNOQARTAPETXNKSNSAYRKAANA	FDNSVIFEAFESNFVYPTPSESANVRIQAQNADE	908
Db	939	tgsittqagskieaktgea	---nvtsatgtlgttsgntvntandf--nitikdgari	992
Qy	909	FASLGTSFEMAPQYNSKDRTEFLDST	INDNGYAFRTORYDLCWSEEP-----NKYGT--	958
Db	993	katgg--avtltatggtlttetsdsdittsg	qntlltakdssiagsiinaanvltntgtlt	1050
Qy	959	-----DEDLRNAIQALHKAGLQVMAD	WPDQIYNLPKREKAVTIVRV	999
Db	1051	tvagskieaasgtlvinakdaqldgaasg	htvvnatna-----ngsgsvtiatts--	1100
Qy	1000	DDRGNVWKDALINNLLVYVNTIG	-----GGEYQYKKY-----GGAFIDKIQ	1033
Db	1101	--srvnitgdltlinglnliskngntvll	kgvdeivkylgpglasvneivleakralekvk	1159
Qy	1040	KLYP-----EIFTKKOYSTGV	ALDPDSQKIT	1064
Db	1160	dlgsdeertlklaklavsavriaebnna	it	1187

RESULT 24
AAU34339
ID AAU34339 standard; Protein; 2434 AA.
XX
XX AAU34339;
XX
DT 14-FEB-2002 (first entry)

QY 1076 ILHRSQSYVLK-----ADGQYVNLGTTTQFLPIQLTG-----EKKQCN 1115
 Db 1831 alngaqnlankgettannlshlnnaqkgdntqvtnapnistvnqvktaeqidqam 1890
 QY 1116 EGFVKNGDNYFYFDLAGNVMKNTFIEDSVGNVFFDDGCKMVENKHFVDVDSYGEKGY 1175
 Db 1891 erling-----lgdk-----dq-----vksqvnftdarp--ekqta 1919
 QY 1176 PFLKNGVSFRGLVOTDNGT 1195
 Db 1920 y--nnavtaeenlingangt 1937
 RESULT 25
 AAB01825
 ID AAB01825 standard; Protein; 1221 AA.
 AC AAB01825;
 DT 11-SEP-2000 (first entry)
 DE Haemophilus influenzae strain Joys mature HMWIA protein, SEQ ID NO:28.
 KW Mature HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
 KW non-typeable Haemophilus influenzae; NTHi; non-encapsulated;
 KW recombinant production; Escherichia coli; antibacterial; vaccine;
 KW human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
 KW detection; diagnosis.
 OS Haemophilus influenzae strain Joys.
 FN WO200020609-A2.
 PD 13-APR-2000.
 PF 07-OCT-1999; 99WO-CA00938.
 PR 07-OCT-1998; 98US-0167568.
 PR 08-DEC-1998; 98US-0206942.
 PA (CONN-) CONNAUGHT LAB LTD.
 PI Loosmore SM, Yang Y, Klein MH;
 XX WPI; 2000-303789/26.
 DR N-PSDB; AAA52176.
 CC Nucleic acid molecule for producing recombinant high molecular weight
 PT proteins of Haemophilus which are used as a vaccine to provide
 PT protection against Haemophilus induced diseases in humans -
 XX Claim 8; Fig 18A-R; 307pp; English.
 CC The invention relates to the recombinant production of Haemophilus
 CC influenzae high molecular weight (HMW) proteins in Escherichia coli. The
 CC expression construct used to effect recombinant expression comprises a
 CC promoter functional in E. coli (e.g., the P7 promoter) operably linked
 CC to a modified hmwABC operon from a non-typeable (non-encapsulated) H.
 CC influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw gene
 CC clusters termed hmwIABC and hmw2ABC. Each hmwABC operon comprises hmwA,
 CC hmwB and hmwC genes. The hmwA genes encode the structural HMW proteins
 CC and the hmwB and hmwC genes encode accessory proteins which are
 CC responsible for post-translational processing and secretion of the HMW
 CC proteins. The modified hmwABC operon used in the expression construct of
 CC the invention contains an A gene modified such that it encodes only the
 CC mature HMWIA. The invention also discloses hmwa genes (AAA52175-A52198)
 CC and HMWA proteins (AAB01824-B01849) from the non-typeable H. influenzae
 CC strains Joys, K1, K21, LCDC2, PMH1, 15 and 12. The nucleic acids and
 CC vectors are used for the production of recombinant H. influenzae HMW
 CC proteins which can be used as vaccines to mediate a humoral or
 CC cell-mediated immune response to provide protection against diseases in
 CC humans caused by H. influenzae (e.g., otitis media, epiglottitis,
 CC pneumonia and tracheobronchitis). The HMW proteins are also useful as

CC antigens in immunoassays for detecting antibodies against Haemophilus,
 CC HMW proteins and/or HMW peptides. The nucleotide sequences encoding the
 CC HMW proteins can be used to isolate and clone hmw genes from other
 CC non-typeable strains of Haemophilus via hybridisation reactions. The
 CC present sequence represents a mature HMWA protein from a non-typeable
 CC strain of H. influenzae.

XX Sequence 1221 AA;

Query Match 3.1%; Score 205.5; DB 21; Length 1221;
 Best Local Similarity 19.9%; Pred. No. 0.00069;
 Matches 237; Conservative 173; Mismatches 464; Indels 317; Gaps 61;

QY 18 NSAVSGPDTTICKLTNDQYQALNGQLQVLLRFSKAADGNPSGDNVTVDQFSKNYATTGGNF 77
 Db 163 naaangitglt-lfnlk-rdlgnnfqinfngtl--nlsagvnlsmvlpkkw----- 211
 QY 78 DYVKVNGNOVEFSGWHATNQSNDKDSQIIVL-VNGKEVKRQLVNDTKEGAAGFNRNDVY 136
 Db 212 dyakfrg----tywvthlnvaegskfnitidsrgddtagtIntpynlingisfnkdtif 267
 QY 137 KVPDAIENSNSGFGQIITLPVTVKNEVQLVHRFSNDVKTGEGNVYDFWSELMPPVKDS- 195
 Db 268 dvk---qngavt-fd--lkapigvnnrnlnyafngnlsvisvggnvtf--kllassta 319
 QY 196 -----FQKNGPLKQPGIQTINGQYYIDPTTQPRKNFLQSGNNWYIPDSOT 244
 Db 320 qtpgvflnskhfnasggsslef-----rtegstkvgfl--nndltina-t 362
 QY 245 GVGNTALELQ-----FAKGTVS-SNEQVRNGNAAYSDDKSIENNVNGYLA----- 289
 Db 363 ggnlslqvegidmigmkgvvaknltfaggnltfg-skkaiteiegnatinnanvtll 421
 QY 290 ---DTWYRPKQILKD-----GTTWTSKETDMPILMWPNLTQAYLYLNTMKQGNL 340
 Db 422 gsdfnhbkptikkdvilnsngnltaggnvlnngnltv--nnganlkaitnftnvvggl 479
 QY 341 LP-----SALPFNDADAPAEALNHYSEIVQONIEKRISGTGNTD-WLRTLMHDFVTN--- 391
 Db 480 fdnkgnsnislarggakfkdnntssl---nl-----ttnsdttvrtflegntnkgad 530
 QY 392 -NPMWNKDSNVNF-SGIQFGGFLKYENSDLTPYANSDYRLGRMPINIKDQ-TYR---G 446
 Db 531 lnifdnkgnaaeiq-----ggnlsqkegnlt--lssd-----klnitqitikkvgn 575
 QY 447 QEFLLANDIDNSNPVQAEQLNWLYLLNFGITIANNDQANFDSYRVDPADPNIDADLMNI 506
 Db 576 kedsdsstannanltiktkeiq-----ltgdlnisgfdkaetakegadlign- 624
 QY 507 AODYPNAAYGMDSDAVSNKHINILEDNHADPEVFNKI-----GNPQLTMDDTIKNSL 559
 Db 625 -----sd-----nnnan-----akkvtnqvkdkisagshvnlsvkts- 662
 QY 560 NHGLSDAPNRMGLDAIVHOSIADRNNSNTENVIPNYSFVR-AHDNNSQDQIONAIRDVT 618
 Db 663 -ngnndaesng-ds---tsltinaknvtvnnnitshtkvtntasenvttkagttinati 717
 QY 619 GKDYTFTFEDEQKIDAYIQDQNSTVK--KYNLYNI-----PASYALLNKKDIP 668
 Db 718 gsvevtaktgdikgiesngnvnitasgdtlnvsntlgqnvtaaaagavtttkgstin 777
 QY 669 RYVYGDLYTDGGYMEHOTRYVYDTLTNLLKSRVYVAGQSQMQTMS-VGGNNNLTISVRY 727
 Db 778 a-----ttgnanittktgeingevksagngvnitasgntlnnvsnitgqnvta--- 827
 QY 728 GKGMATATDTCTDEFTQIGIGVVVSVNTNMLKLGVDKVKVVLHMGAAHKKQYRAAVLTFTD 787
 Db 828 nsaglttie-gstlnatigdanittgtn-----ingkvesssg-----svtliatg 873
 QY 788 GVINYSQCAPVAMTDENGDLYLSHNLVYNGKEADTAVOGYANPDVSGYLA-VWTPV 846
 Db 874 qtlavngisgdtvtitadkgtittgtss-kingtksvttssq---sgdisgtisgntsvs 929

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 12, 2002, 07:52:43 ; Search time 16.41 Seconds
(without alignments)
1862.061 Million cell updates/sec

Title: US-09-995-749A-2_COPY_531_1781

Perfect score: 6641

Sequence: 1 MSRTAPDVAAYPSLYNSA.....SDAEXPTSTDVGKMLDNKL 1251

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 2442594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

- 1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/2/iaa/6CTUS_COMB.pep.*
- 6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2446	36.8	1430	3	US-09-008-172-2	Sequence 2, Appli
2	2446	36.8	1430	4	US-09-210-361-6	Sequence 6, Appli
3	2347.5	35.3	1375	4	US-09-210-361-4	Sequence 4, Appli
4	2276	34.3	1475	3	US-09-007-999-2	Sequence 2, Appli
5	2276	34.3	1475	4	US-09-210-361-2	Sequence 2, Appli
6	2175.5	32.8	1577	2	US-08-793-834-2	Sequence 2, Appli
7	326.5	4.9	349	3	US-09-009-620-2	Sequence 2, Appli
8	245.5	3.7	2314	4	US-09-268-347-49	Sequence 49, Appli
9	222	3.3	2366	1	US-08-480-604A-10	Sequence 10, Appli
10	222	3.3	2366	2	US-08-405-436A-10	Sequence 10, Appli
11	222	3.3	2366	4	US-08-913-136-10	Sequence 10, Appli
12	209	3.1	2710	1	US-08-480-604A-6	Sequence 6, Appli
13	209	3.1	2710	2	US-08-405-436A-6	Sequence 6, Appli
14	209	3.1	2710	4	US-08-913-136-6	Sequence 6, Appli
15	208	3.1	2353	4	US-09-377-153-33	Sequence 33, Appli
16	208	3.1	2353	4	US-08-913-942-4	Sequence 4, Appli
17	208	3.1	2353	4	US-09-669-974-33	Sequence 33, Appli
18	206	3.1	1296	3	US-08-470-260-3	Sequence 3, Appli
19	206	3.1	1296	3	US-08-471-491-3	Sequence 3, Appli
20	206	3.1	1296	4	US-08-466-662-3	Sequence 3, Appli
21	203.5	3.1	2354	4	US-09-268-347-47	Sequence 47, Appli
22	201	3.0	1403	2	US-08-387-942C-3	Sequence 3, Appli
23	200.5	3.0	1938	4	US-09-514-302-2	Sequence 2, Appli
24	200.5	3.0	2123	4	US-08-968-685A-10	Sequence 10, Appli
25	200.5	3.0	2411	4	US-09-268-347-36	Sequence 36, Appli
26	197.5	3.0	1612	1	US-08-169-927-2	Sequence 2, Appli
27	195	2.9	1541	4	US-08-296-791-3	Sequence 3, Appli

28	195	2.9	1541	5	PCT-US95-10661A-3	Sequence 3, Appli
29	192.5	2.9	1529	2	US-08-728-470-10	Sequence 10, Appli
30	192.5	2.9	1529	4	US-08-719-641-10	Sequence 10, Appli
31	192.5	2.9	1600	2	US-08-617-697-10	Sequence 2, Appli
32	190	2.9	1287	1	US-08-200-232-2	Sequence 2, Appli
33	190	2.9	1287	5	PCT-US95-02219-2	Sequence 2, Appli
34	190	2.9	1287	5	PCT-US95-02219A-2	Sequence 2, Appli
35	189.5	2.9	631	3	US-08-814-052-8	Sequence 8, Appli
36	189.5	2.9	631	3	US-08-812-829-8	Sequence 8, Appli
37	187.5	2.8	750	3	US-08-814-052-2	Sequence 2, Appli
38	187.5	2.8	750	3	US-08-812-829-2	Sequence 2, Appli
39	187.5	2.8	1104	4	US-09-268-347-28	Sequence 28, Appli
40	187.5	2.8	1104	4	US-09-268-347-34	Sequence 34, Appli
41	187	2.8	1507	6	5268270-2	Patent No. 5268270
42	186.5	2.8	1912	1	US-08-409-995-4	Sequence 4, Appli
43	186.5	2.8	1912	3	US-08-685-467-4	Sequence 4, Appli
44	185.5	2.8	485	4	US-09-291-023A-18	Sequence 18, Appli
45	184	2.8	2285	4	US-09-308-375-2	Sequence 2, Appli
46	183.5	2.8	1536	1	US-08-038-682-2	Sequence 2, Appli
47	183.5	2.8	1536	1	US-08-302-832-2	Sequence 2, Appli
48	183.5	2.8	1536	2	US-08-530-198-2	Sequence 2, Appli
49	183.5	2.8	1536	2	US-08-459-880-2	Sequence 2, Appli
50	183.5	2.8	1536	2	US-08-728-470-2	Sequence 2, Appli
51	183.5	2.8	1536	2	US-08-617-697-2	Sequence 2, Appli
52	183.5	2.8	1536	4	US-08-719-641-2	Sequence 2, Appli
53	182.5	2.7	484	6	5171673-8	Patent No. 5171673
54	182.5	2.7	487	6	5171673-6	Patent No. 5171673
55	181	2.7	1183	2	US-08-447-031A-2	Sequence 2, Appli
56	180.5	2.7	1702	4	US-08-296-791-5	Sequence 5, Appli
57	180.5	2.7	1702	5	PCT-US95-10661A-5	Sequence 5, Appli
58	180	2.7	560	3	US-08-814-052-6	Sequence 6, Appli
59	180	2.7	560	3	US-08-812-829-6	Sequence 6, Appli
60	177.5	2.7	483	1	US-08-468-700-34	Sequence 34, Appli
61	177.5	2.7	483	1	US-08-645-371-2	Sequence 2, Appli
62	177.5	2.7	483	2	US-08-468-220-32	Sequence 32, Appli
63	177.5	2.7	483	2	US-08-468-698-32	Sequence 32, Appli
64	177.5	2.7	483	3	US-08-704-706A-34	Sequence 34, Appli
65	177.5	2.7	483	3	US-08-890-383-3	Sequence 3, Appli
66	177.5	2.7	483	3	US-08-914-679A-3	Sequence 3, Appli
67	177.5	2.7	483	4	US-09-182-859-2	Sequence 2, Appli
68	177.5	2.7	483	4	US-09-170-670-4	Sequence 4, Appli
69	177.5	2.7	483	4	US-09-193-068-4	Sequence 4, Appli
70	177.5	2.7	483	4	US-09-183-412-4	Sequence 4, Appli
71	177.5	2.7	483	4	US-08-985-659-35	Sequence 35, Appli
72	177.5	2.7	483	4	US-09-264-097-2	Sequence 2, Appli
73	177.5	2.7	483	4	US-08-194-664A-32	Sequence 32, Appli
74	177.5	2.7	483	4	US-09-291-023A-21	Sequence 21, Appli
75	177.5	2.7	483	5	PCT-US94-01553A-32	Sequence 32, Appli
76	177.5	2.7	483	5	PCT-US95-10426-32	Sequence 32, Appli
77	177.5	2.7	487	2	US-08-468-220-37	Sequence 37, Appli
78	177.5	2.7	487	2	US-08-468-698-37	Sequence 37, Appli
79	177.5	2.7	487	4	US-08-194-664A-37	Sequence 37, Appli
80	177.5	2.7	487	5	PCT-US94-01553A-37	Sequence 37, Appli
81	177.5	2.7	487	5	PCT-US95-10426-37	Sequence 37, Appli
82	177.5	2.7	512	1	US-07-623-953-3	Sequence 3, Appli
83	177.5	2.7	512	1	US-07-623-953-5	Sequence 5, Appli
84	177.5	2.7	512	1	US-08-730-899-2	Sequence 2, Appli
85	177.5	2.7	512	1	US-08-459-610-2	Sequence 2, Appli
86	177.5	2.7	512	2	US-08-343-804-2	Sequence 2, Appli
87	177.5	2.7	512	2	US-08-687-399-2	Sequence 2, Appli
88	177.5	2.7	512	3	US-08-600-908A-2	Sequence 2, Appli
89	177.5	2.7	512	3	US-08-683-838A-2	Sequence 2, Appli
90	177.5	2.7	512	4	US-08-985-659-34	Sequence 34, Appli
91	177.5	2.7	2048	4	US-09-288-347-48	Sequence 48, Appli
92	174.5	2.6	483	2	US-08-468-220-36	Sequence 36, Appli
93	174.5	2.6	483	2	US-08-468-698-36	Sequence 36, Appli
94	174.5	2.6	483	4	US-08-194-664A-36	Sequence 36, Appli
95	174.5	2.6	483	5	PCT-US94-01553A-36	Sequence 36, Appli
96	174.5	2.6	483	5	PCT-US95-10426-36	Sequence 36, Appli
97	174.5	2.6	1545	4	US-08-296-791-4	Sequence 4, Appli
98	174.5	2.6	1545	5	PCT-US95-10661A-4	Sequence 4, Appli
99	173	2.6	1338	2	US-08-728-470-9	Sequence 9, Appli
100	173	2.6	1338	4	US-08-719-641-9	Sequence 9, Appli

ALIGNMENTS

RESULT 1
US-09-008-172-2
; Sequence 2, Application US/09008172
; Patent No. 6127602
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starch and
; TITLE OF INVENTION: Latexes in Paper Manufacture
; FILE REFERENCE: 0358D
; CURRENT APPLICATION NUMBER: US/09/008,172
; CURRENT FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1430
; TYPE: PRT
; ORGANISM: streptococcus mutans
US-09-008-172-2

Query Match 36.8%; Score 2446; DB 3; Length 1430;
Best Local Similarity 46.7%; Pred. No. 3.3e-167;
Matches 508; Conservative 170; Mismatches 310; Indels 100; Gaps 25;

QY 208 LQTINGQYYIDPTTGPQRKNFLQSGNNWYIFSDTGVGTNALELQFAKTVSSNEQYR 267
DB 174 VKQIDGKYYTIG-SDGPKPKFALTVNKNVYFDKNTGALTDTSQYQKGLTKLNNDYT 232

QY 268 NGNAYSYDDKSIENVNGVYLTADTWYRKQILKDCGTTWTDKSETDMRPIKAWMPNTLTQ 327
DB 233 PHNQIVFENFSLIEDINQYVADSWYRKDILKNGKTYTASESDLRPLLSWPDKQYQ 292

QY 328 AYLYNMQKHG-----NLLPSALPFNADAPAEHLNHYSEIVQVQNIKRISETGNTDMLR 382
DB 293 IAYLYNMQKGLGTGEN-----YTADSSQESLMLAAQTQVQVKIETKISQYQQTQTLR 344

QY 383 TLMHDFVTNNPMMKDSNVNFSGIQ--FOGFLKYENSDLTPYANSYRLLGRMPIN-- 438
DB 345 DIINSEFVKTPQNNNSQTESDTSAGEKHLQGGALYYSNDSKTPAYANSYRLLNRPPTSQT 404

QY 439 -----IKDQYRQGBFLANDIDNSNPVQAEQLNKLNYLLNFGTITANNQANFDSYRV 493
DB 405 GKPKYFEDNSSGGYDFLLANDIDNSNPVQAEQLNKLNYLNNYSGIVANDPEANFDGVRV 464

QY 494 DAPDNIDADLMIAQDYFNAYGMD-SDAVSKNHINILEDHWHADPEYFNKIGNPQLPHD 552
DB 465 DAVDNVNADLLQIASDYLAHYGVDSKERNAINHLISLEANSNDPQYNKRTKGAQLPID 524

QY 553 DPKNSLHGLS-----DATN----RWGLDAVHOSLADRENNSTENVVIPNYSFVRAHD 603
DB 525 NKRLSLLYALRPLEKDSAKNETRSGLEPVITSLNRSSEGNSEMANYPFIRAH 584

QY 604 NNSQDQIQNAIR-DVTGK-DYHTFTFEQKGDIDAYIQDQNSTVKKYNLYNIPASYAILL 661
DB 585 SEVQTVIAKIIKRAQINPKTDGLTFLDELKQAFKIYNEQMRQAKKKYQTSNPTAYALML 644

QY 662 TNKDTIPRYVGLYTDGQYMEHQYTRYDTLTNLKSRVKYVAGQSQMOTMSVSGNNR- 720
DB 645 SNKDSITRLYGDYSDDGQYMATKSPYYDAITLLKARIKTAAGGQDMKITIYVPEGKSH 704

QY 721 -----ILTSYRYGKGMATDTGTDEFTGTGIGVWVSNLNLKLVNDKVVVLHMGAAH 773
DB 705 MDWDYGVGVTSTYRYGANEADTQSEATKTQGMVAVTSTNNFSLNQNQDKVIVNMGAAH 764

QY 774 KMQYRAAVLTITTDGYINTSDQAP--VAMTDENGDIYLSHNLVNGKEADTAVQGY 831
DB 774 KMQYRAAVLTITTDGYINTSDQAP--VAMTDENGDIYLSHNLVNGKEADTAVQGY 831

DB 765 KQGEYRPLLLTTTKDGLTSYTSDAARAKSLYRKTNDKGELYDASD-----IQGY 812

QY 832 ANPDVSGYLAVWYPVPGASDNQDKARTAPSTKNSGNSGAYRTNAAFDNSVIFEPSNFVYTP 891
DB 813 LNPQVSGYLAVWYPVPGASDNQDKVRVAASNKANATQGVYESSALDSQLIYEGFSNFQDFV 872

QY 892 TKESERANVRIAQNAOFFASIGFTSFEMAPQYNSSDRTFELDSTIDNGVAFTRDYDLGMS 951
DB 873 TKDSYTNKKIARVQLFKSWGVTSFEMAPQYVSSDSGLDSIIQNGYAFEDRYDLAMS 932

QY 952 EPNKYGTDEDLRNAIQALHAKGLQVWADWYPDQIYNLPKGEVATVTRVDDRGVWVKDAII 1011
DB 933 KNNKYSGQDMINAVKALHKSIGIYADWYDQIYNLPKGEVATVTRVDDRGVWVKDAII 992

QY 1012 NNNLYVNTTGGG-EYQKYGGAFLDKLQKLYPEIFETKKQVSTGVALDPSQKITEWSAKY 1070
DB 993 KNTLYAANTKSNKDYQAKYGGAFSLSELAAYPSIFNRTQISNGKKIDPSEKITTAKKY 1052

QY 1071 FNGTNILHRSGVYVKADGG-QYVNL-GITTKQFLPQLGKKGKOGNEGVKNGDGN-YY 1127
DB 1053 FNGTNILGRGVYVLDKDNASDKYFELKGNQY--YLPKQMT--NKEASTGFV--NDGNGWT 1106

QY 1128 FYDLAGNWKVNTFIETDSVGNWYFPDQDKMVENKHFYDVSYGEGKTYFFELKNGVSFRGG 1187
DB 1107 FYSTSGYQAKNSFVQDAKNWYFDNNGHMYGLQQLN---GE--VQYFSLNGVQLRES 1160

QY 1188 LVQT-----DNGIYFFDNYGK-----MYRNQITINAGAMYITLDEN 1222
DB 1161 FLENADSKNYFHLGNRYNGYYSFNDMSKRYFDASGVNAVLKTINGTQYF--DQD 1218

QY 1223 GKLIKASY 1230
DB 1219 GYQVKGAW 1226

RESULT 2
US-09-210-361-6
; Sequence 6, Application US/09210361
; Patent No. 6284479
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starches and
; TITLE OF INVENTION: Latexes in Paper Manufacture
; FILE REFERENCE: 0357CR
; CURRENT APPLICATION NUMBER: US/09/210,361
; CURRENT FILING DATE: 1998-12-11
; EARLIER APPLICATION NUMBER: 09/007,999
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/009,620
; EARLIER FILING DATE: 1998-01-20
; EARLIER APPLICATION NUMBER: 08/485,243
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/008,172
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 1430
; TYPE: PRT
; ORGANISM: streptococcus mutans
US-09-210-361-6

Query Match 36.8%; Score 2446; DB 4; Length 1430;
Best Local Similarity 46.7%; Pred. No. 3.3e-167;
Matches 508; Conservative 170; Mismatches 310; Indels 100; Gaps 25;

QY 208 LQTINGQYYIDPTTGPQRKNFLQSGNNWYIFSDTGVGTNALELQFAKTVSSNEQYR 267
DB 208 LQTINGQYYIDPTTGPQRKNFLQSGNNWYIFSDTGVGTNALELQFAKTVSSNEQYR 267

Db 174 VQIDQKYYIG-SDGQPKKFNALTVNNKVLFDKNTGALTDTTQYQFKQGLTKLNNDYF 232
QY 268 NGAAYSDDKSIEWNGYLAADWYRKPQILKDGTTWDSKETDMRPILAVWNPNTLTQ 327
Db 233 PHNQIVNEFTSLETIDNYTADSYRKPDLKNGKTWTASSEDRLPRLSWPDKQTQ 292
QY 328 AYLVNMQHG-----NLLPSALPFNADAPAEUNHYSEIVQOENIKRISSETGNTDWR 382
Db 293 IAYLVNMQGLTGEN-----YTADSSQESINLAQTVOVKIETKISQTOQTQWLR 344
QY 383 TLMHDFVTNNPMWKNDSNVNFGIO--FQGGFLKYENSDLTPPYANSRYLRGLRMPIN-- 438
Db 345 DIINSFVKTPQNNWQSTESDAGEKHQGLGALLYSNDKTAYANSRYLRNRPSTQ 404
QY 439 -----IKDQTYGQBELANDNDNPNVYQABOLNWLNYLLNFGFTANNQANEDSVRV 493
Db 405 GKPKYFEDSSGQYDFLLANDNDNPNVYQABOLNWLNYLLNFGFTANNQANEDSVRV 464
QY 494 DAPNDIDADLMIADYFNAAYGMD--SDAVSNKHINILEDWHDADPEYFNKIGNPQLTMD 552
Db 465 DAVDNWADLLQIASDYLKAHYGVDSKSEKNAIHLISLEWSDNDPQYNKDTKGAQLPID 524
QY 553 DTIKNSLNHGLS-----DATN-----RWGLDAIVHQSADRENNTENVPYIPSVFRAHD 603
Db 525 NKLRSLYALTRPLEKADSNKNEIRSGLEPVITNSLNRSAEGKNSERMANYIFIRAH 584
QY 604 NNSQOIQNAIR-DWTGK-DYHTFTFEDEQKIDAYIODNSTVKKYNYLYNPASVAILL 661
Db 585 SEVOTVIKIAIKQINPKTDGTFTELDLQAFKYNEDMRQAKKKYQTSNIPYATYALML 644
QY 662 TNKDRIPRYVYDGLYDQGYMEHQTTRYDITNLLKSRVYVAGQSMQTSVGGNNH-- 720
Db 645 SNKDSITRLYGDWSDDQGYWATKSPYIDA.DTLKARIKAAAGQDKIYVEGDKSH 704
QY 721 -----ILTSVRYGKAMTADTGTDETRTQIGIGYVVSNTPLNKLGVNDKYVLHMGAAH 773
Db 705 MDWDTYGLTSLVRYGTGANEATDOGEATKQGMVAVITSNPSLKNQNDKVIYVWGAH 764
QY 774 KQOYRAAVLITTDGVTNYSQDAP--VAMDENDGLDYLSSHNLVNGKEADTAVQGY 831
Db 765 KNOEYRPLLTTKDGLTSTYSDAAKSLYRKNTDNGELVFDASD-----IQGY 812
QY 832 ANPDVSGYLAVWPVPGASQNDARTAPSTSEKNSGNSAYRTNAAFDSNVIFAEAFSNFYTP 891
Db 813 LNPQYSGYLAVWPVPGASQNDQVRYAASNKANATQVYESSALDSQLIYEGFSNFQDVF 872
QY 892 TYSEERANRYAQNADDFASLGETSFEMAPQYNSSKDRFTLSDTDNGYAFDRDRLGMS 951
Db 873 TRKDSYTNKKIRQNVQLFKSWGVSFEMAPQVVSSEDDGSLDLSIIQNGYAFEDRYDLAMS 932
QY 952 EPNKYGTDEDLNAIQALHKAQVMAWVPOQIYNLPQKEVATVTRVYDRGNVWKDAII 1011
Db 933 KNNKYSQODMNAKALHKSQIOWADWVPOIYNLPQKEVATVTRVYDRGNVWKDAII 992
QY 1012 NNNLVVNVYIGGG--EYQKYGAPLQKQKLYPEIFTKKQVSTGVAIDPSQKITEWSAKY 1070
Db 993 KNTLYAANTKSGDKYQAKYGAFLSELAAYPSIFNRQIISNGKKIDPSEKITAWKAY 1052
QY 1071 FNGTNIHLRGSGYVLKADGG-QYVNL-GTTTQKQFLPIQLTGEKKQGNQEGFYKNGDN-YY 1127
Db 1053 FNGTNIHLRGSGYVLKADGG-QYVNL-GTTTQKQFLPIQLTGEKKQGNQEGFYKNGDN-YY 1106
QY 1128 FYDLAGNMVKNFTEDSVGNTFFDQDQKVENKHPVDVDSYGEKGYTFFLKNGYSFRGG 1187
Db 1107 FYDLSGQYKNSFVQAKKNWYFDNNGHMYGLQQLN-----GE--VQYFLSNGVQLRES 1160
QY 1188 LVQNT-----DNGTYTYDNTYK-----MYRQNTINAGAMIYTLDEN 1222
Db 1161 FENADGSKNYFCHLGNRYSGNYISFDNDSKRWYFDASGVMAVGLTNGTYQF--DQD 1218
QY 1223 GKLKASY 1230
Db 1219 GYQVKGAW 1226

RESULT 3
US-09-210-361-4
; Sequence 4, Application US/09210361
; Patent No. 6284479
; GENERAL INFORMATION:
; APPLICANT: Nicholls, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starches and
; TITLE OF INVENTION: Latexes in Paper Manufacture
; FILE REFERENCE: 0357CR
; CURRENT APPLICATION NUMBER: US/09/210,361
; CURRENT FILING DATE: 1998-12-11
; EARLIER APPLICATION NUMBER: 09/007,999
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/009,620
; EARLIER FILING DATE: 1998-01-20
; EARLIER APPLICATION NUMBER: 08/485,243
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/008,172
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 1375
; TYPE: PRT
; ORGANISM: streptococcus mutans
US-09-210-361-4

Query Match 35.3%; Score 2347.5; DB 4; Length 1375;
Best Local Similarity 41.3%; Pred.No.3.7e-160;
Matches 538; Conservative 181; Mismatches 408; Indels 175; Gaps 35;
QY 19 SAVSGFDFTIKLNDQYQALNGQLVLLRFKADG--NPSGDNVTVDQFSKNYATGNG 76
Db 29 TSLSSGLVKADSTDDROQAVTESQASLVTTSEAAKETLTATDTSTATSATQPTATVDN 88
QY 77 FDYVKNVNGQVEFSGHWATNQSNDKDSQWIIIVNGKEVKQLVNDTKREGAAGFNRNDV 136
Db 89 VS-----TTNQS-----TNTANTA-----NPFV 107
QY 137 KVNPALENSWSGFGIITLPPVYVKNVQVLRHPSNDVKTGEGNVY-----DFW 186
Db 108 KPTITSEQAKTDNSDKIITTSKAVNR-----LTATGKFVPANNNTAHPKTVT 154
QY 187 SELMPVKDSFQKNGPLKQ-----FGLQTINGOQVYIDPTGQPKKELL 231
Db 155 DKIVPIKFI-----GKLQPSLSQDDTAALGNVKNIRKVGKYYY--KEDGTQKNTAL 209
QY 232 QSGNNMIYDSDTGYGTNALELQFAKGTVSSNE---QYRNGNAAYSYDDKSTENWNGYLT 288
Db 210 NINGKTFEDEFEGALSNNLPSK--KGNITNDNTNSAQYNOYVSTDVAFEHVHILT 267
QY 289 ADTWRPQILKDGTTWDSKETDMRPILAVWNPNTLQAYLYNPMKHQNLPLSALPFF 348
Db 268 ABSWTRPKYILKDGKWTQSTEKDFRPLMTWVDPQETQRYVYVYVYVYVYVYVYVYVY 323
QY 349 NADADPAELNHYSEIVQOENIKRISSETGNTDWRITLMDHFTVNNPMWKNDSNVNFSQ 408
Db 324 WTATSEQLNLAQIQIKIEKIPAEKNTNWLROTISAFVKTQSANNWSEKFPDDHLQ 383
QY 409 FQGGFLKYENSDLTPYANSRYLRGLRMPIN-----IKDQTYGQBELANDNDN 459
Db 384 -KGALLYSNNSKLTQANSNYRILNRTPTNQTGKKDPRYTADRTIGGYEFFLLANDVDNS 442
QY 460 FVYQABOLNWLNYLLNFGFTANNQANEDSVRVDPADNIDALMIADYFNAAYGMD 518
Db 443 PVYQABOLNWLHFLNFGNIYANDPDANFDSIRVDVNDVADNVDADLQIAGDYLKAAKGIH 502


```

Db      956  AGSQIKNTVTVVDGKSGKDDQAKYGGAPLEBJAQKAYPELFARKQISTGVPMPSVAKIKQ 1015
QY      1066  WSAKYFNGTNILHRGSGYVLKADG-QGYINLGTITK-QFLPQLTGTGKKKQGNEGFYKGN 1123
Db      1016  WSAKYFNGTNILHRGAGYVLKQATNTYFENISDNKEINFLPKTLNQDSQ--VGFSYDGG 1073
QY      1124  GNTYFYDLAGNMVKNFTFEDSGVGNWFFDQDQCKMYENKHFYVDVDSYGEKGTGYFFLKNGVS 1183
Db      1074  G-VVYVYSTGYQAKNTFISEG-DKWVYFDNNGYVMVTAQSIN-----GVNYFYFLSNGLQ 1125
QY      1184  FRGLVQTDNGTY-FEDNKGKVMRN-----QTINACAM-----IYTLDEN 1222
Db      1126  LRDAILLKNEGTAYVGNDCRRYENGTYQFMGVMRHFNGEMSVGLTVIDGQVQYFDEM 1185
QY      1223  GKLIKASYNSDAEYPTSTDGKM 1245
Db      1186  G-----YQAKGKFYTAD-GKI 1201

RESULT 5
US-09-210-361-2
; Sequence 2, Application US/09210361
; Patent No. 6284479
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starches and
; TITLE OF INVENTION: Latexes in Paper Manufacture
; FILE REFERENCE: 0357CR
; CURRENT APPLICATION NUMBER: US/09/210,361
; CURRENT FILING DATE: 1998-12-11
; EARLIER APPLICATION NUMBER: 09/007,999
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/009,620
; EARLIER FILING DATE: 1998-01-20
; EARLIER APPLICATION NUMBER: 08/485,243
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/008,172
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-210-361-2

```

```

Query Match      34.3%; Score 2276; DB 4; Length 1475;
Best Local Similarity 42.4%; Pred. No. 5.7e-155;
Matches 518; Conservative 189; Mismatches 382; Indels 140; Gaps 37;
```

QY	107	I LVNGEVRQLVND-----TREGAAGFNDRDYKYNPALENSSM	147
	:	: : : :	:
Dd	35	L VKADSNESKSQISDNGNTSVYTANESNVITEATSKQEAASSQT--HTVTTSSTSYV	92
	:	: : : :	:
QY	148	S GFQGIITLPVTKNENVLHFRHSNDVKTEGEGNYDFWSEL-----MPVKD	194
	:	: : : :	:
Dd	93	V NPXEVVSNPYTV-GETASNGEKIQNOTT-----VDKTSEAAANNISKQTEAUTDVID	146
	:	: : : :	:
QY	195	S FQKGNGPL--KQFGLQTINGQQYYIDPTQCPKRNFLQSGNNWIYFD-----SDTGVG	247
	:	: : : :	:
Dd	147	DSNAANLQILEKLPNKVEIDKKYYYD--NNGKVTFNFTLIADGXILHFDEGTAYDTSID	205
	:	: : : :	:
QY	248	TNALQLQFAKGVTSVSSNEQRYNGNAAYSDDKS IENVGYLFADWYRPKQLDKGTTWTD	307
	:	: : : :	:
Dd	206	TVNKDIV---TTRSN-LYKKYNQYDESAGEFWHDYLTABSWYRKYILDKGKWTFQ	260
	:	: : : :	:
QY	308	SKETDMRPILMVWPNTLTQAYYINVMKHGNLPSALPFFVNADDPACLNLHSEIVQN	367
	:	: : : :	:

261	Db	STBKDFRPLMTWTPDQETQROYVNYMNAOLGINKT-----YDTSNQIQLNAAATIQAK	316
368	QY	IERRISETGNTDLRLTMHDFVTNNPMNKSESNVFSGIOFGGFLKYEN-SDLTPYAN	426
317	Db	IEAKITTLKNTDLWRTISAFVKYTOSSANSSSEK-PPFD-HLQNGAVLYDNEGKLTIPYAN	374
427	QY	SDVRLGRMPIN-----IKDQTYRGQEFLLANDIDNSNPVYVQABQLMWLYLLNFG	477
375	Db	SNTRIILNRTPTNQTGKKDPRTYADNTIGGYEFLLANDVDSNPVYVQABQLMWLHFLNFG	434
478	QY	TITANNDQANESVRDAPONTDADLMIATQDYFNAAVYCM-DSDAVSNKHNLINILEDNHA	536
435	Db	NIYANDPDANFISIRVDVANDVNDQADLLQTAGDYLKAAGYIHKNDKAANDHLSILEAWSDN	494
537	QY	DPEYFNKIGNPOLMTDITKNSLNHGLSDATN-RWGLDAIVHQSLADRENNSNENWYIPN	595
495	Db	DTPLYLHDGDNMINMDNKLRLSLLPSLAKPLNORSGMPLITNSLVNRTDDNATAVPS	554
596	QY	YSFVRAHDNNSQDQIONATR-----DVYTKDYHVTTFDDBQKGDIDAYIQDQNSPVKKYNL	650
555	Db	YSFIRAHDSVODLADIILKABINPNVVG--YSETMEIKKAFEYINKOLLATEKKYTH	611
651	QY	YNTPASVAILLNKDTPIRVYTGDLYTDSGQYMEBQTRYDYTLNLKLSRVKYVYAGQSM	710
612	Db	YNTALSVALLLNKSSPVVYIGDMFTDQGYMAKNTINYEAIEFTLLKARILKYVSGSQGA	671
711	QY	QTVMSGVGNMILTSVRYGKAMTATDGTDETRTOGIVVYSNTPNLKLVNDKXVFLRMG	770
672	Db	RNQOV-GENSEILTSVRYGKALKATDGTGTRTTSVAVIEGNNPSLRLKASDRVYVNMG	730
771	QY	AAHKNQYRAAVLTITDGVINTYSQGAH--VAMTDNGDLYLSHNLVUNGKEBATAY	828
731	Db	AAHKNQAYRPLLTDTTNGIKAYHSQDEAGLVRYVYNDRGELIFTA-----AD--I	778
829	QY	QGVANPDVSGYLVAWVPVGCAS--DNODARTAPSTEKNSGNSAYRTNAAFDNSVIFAFPSN	886
779	Db	KGVANQVSYGLVWVPVGCALIKFALFLA---RPHQOMASVYHQNAALDSRVMEFGFSN	835
887	QY	FVYTPPKESERANVRIAQNAFDFASLGFTSFEMAPQYNSKKDRFLDSTIDNGYAFDTRY	946
836	Db	FOAFATKKEEYTVVITAKNVDPFAENGVTDFEMAPQYVSYSDGSFLDSVIQNGYAFDTRY	895
947	QY	DLGMSFPNKYGTDEDLRNAIQALHAKAGLYOVHADVPDQIYNLPKEVATYVTRVDRGNW	1006
896	Db	DLGISPNKYGTADDLVKAIKALHSGKIKVADWVPDQYAFPEKEVYVTRVDRKYGTPV	955
1007	QY	KDAIINNLYVYN-TIGSGEYQKTKGGAFDLKLOKLYPEITFKQVSTGVAIDPSQKITE	1065
956	Db	AGSQIKNTLYVVDGKSSGKQDQAKYGGAFLELOAKYFELFARKQISTGVMPDPSVVIKQ	1015
1066	QY	WSAKYFNGNLIHRSGYVLLKADG-QQYVNLGTTTK-QFLPIQLTGEKKQNGEYFKGND	1123
1016	Db	WSAKYFNGNLIHRGAGYVLLKADQATNYFNISDNKEINFPLKTLNLQDSQ--VGFSDYDK	1073
1124	QY	GNVYFYDLAGNMVKNFTFIEDSVGNMYEFDQDQKMYENKHFVDVDSYGEKGYFYFLKNGVS	1183
1074	Db	G-VYVYSTSGYQAKNFTFISEG-DKWYFYEDNNYVMTGAOSIN-----GVNYFYFLSGLQ	1125
1184	QY	FRGGLVQTDNGTY-YFDNTYGKVMRN-----QTINAGAM-----TYTLDN	1222
1126	Db	LRDAILKNEDGTVAIYVNDGRVYENGYQFMGVSQVWRHFNNGEMSGLTVIDGQVOYFDEM	1185
1223	QY	GKLIKASVNSDAEYPTSTOVGMK	1245
1186	Db	G-----YQAKGFVTTAD-GRI	1201

RESULT 6
US-08-793-824-2
; Sequence 2, Application US/08793824
; Patent No. 5981838
; GENERAL INFORMATION:
; APPLICANT: Simpson, Christine Lynn

APPLICANT: Giffard, Philip Morrison
 APPLICANT: Jacques, Nicholas Anthony
 TITLE OF INVENTION: Genetic Manipulation of Plants to
 TITLE OF INVENTION: Increase Stored Carbohydrates
 NUMBER OF SEQUENCES: 2
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Griffith Hack & Co
 STREET: Level 8, 168 Walker Street
 CITY: No. 5981838th Sydney
 STATE: New South Wales
 COUNTRY: Australia
 ZIP: 2060
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/793,824
 FILING DATE:
 CLASSIFICATION: 800
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: AU 947643
 FILING DATE: 24-AUG-1994
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 61 2 9957 5944
 TELEFAX: 61 2 957 6288
 TELEX: 26547
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1577 amino acids
 TYPE: amino acid
 STRANDEDNESS:
 TOPOLOGY: not relevant
 MOLECULE TYPE: protein
 ORIGINAL SOURCE:
 ORGANISM: Streptococcus salivarius
 US-08-793-824-2

Query Match 32.8%; Score 2175.5; DB 2; Length 1577;

Best Local Similarity 40.4%; Pred. No. 1.1e-147;

Matches 518; Conservative 189; Mismatches 415; Indels 159; Gaps 38;

QY 19 SAVSGFDITIKLTN---DQYQALNGOLOVLLRFSKAADGNPSGDN-TVDQPSKNVATG 74
 DB 99 SAVSQEATAQATSPVASEVAVSQTQ-----SSQETQTEQVSQQTSP- 145
 QY 75 GNFYVYKNGNOVEFSGWHATNOSKXDSQWIIYLVN-GKEYKROLVNDTKEGAAGFNRN 133
 DB 146 -----QVAGQTSAGQTSFVTEQARPR-----VLTAAPAATAAADSTIRINA--NRN 191
 QY 134 DYYKVNPALENSMSGFQGIITLPYTVK-----NENVQLVHRESNDVKNGCNY 182
 DB 192 ----TNIIITAGTTPNVIITGTPNPNVITSPNGTRPNVITQPNQPNK----- 241
 QY 183 VDFSELMPVKDSF-QKNGPLK-----QFGIQTNGQOYIIDPTGQPKNPL 230
 DB 242 -----PVQFSQPSQPNKVPQPNQPSLDYKPVASNLKTIQKQYVE--NGVYKKNAA 291
 QY 231 LOSGNWIIYFSDGTGVGNALBQAKGTVSSNEQYRNGNAAYSYDDKSIENVNGYLAD 290
 DB 292 IELDGLRYFED-ETGAMVDQSKPLYPADAIIPNNSIYAVYNOAYDTSKSFELHDLNFLAD 350
 QY 291 TWYRPOILKDCGTTTDSKETDWRPLMWMWENTLTQAYILNMYKHGNLPLSALFPFNA 350
 DB 351 SWIRPQILKDKGNWTASTEKDYRLLMTWPDKVTQVNLNYSQCG-----FGNKTTT 406
 QY 351 DADPAELNHYSEIVQONIKRISSETGNTDMLTLMEDFTVNNPWNKDSSE-VNPFSGIQF 409
 DB 407 DMMSYDLAAAEVVGRIEERICRGNTWHLRLMSDIKTIQPGWNSSEDMILVKGKDLH 466
 QY 410 QGGFLKIENSIDLTPYANSYRLLGRMPINIK-----DQTVRGQEFLLANDIDNSNPV 462

DB 467 QGGALTFNNSATSHANSDFRLMNRITFNQGTGRKYHIDRSNGGTELLLANDIDNSPAV 526
 QY 463 QAEQLNLWLYLLNFGTITANNDQANFDSVRVDAPDNIDADLNIAQDYFNAAYGK-DSDA 521
 DB 527 QAEQLNLWELHYTNIGSILGNOPSAFDCGVRIDAVDNVDADLLQIASDYFKEKYVADNEA 586
 QY 522 VSNKHINLEDNHADPEYFNKIGNPQLTMDDTIKNS-LNHGLSDATNRWGLDAIVHOSL 580
 DB 587 NAIHLSTLEAWSYNHQNKDITGAQLSIDNPLRETLTLTFLRKSNYRGSLEIVTNSL 646
 QY 581 ADRENSTENVIPNYSFVRAHDNNSQDQIONAI-RDVTGK-DYHTTFEEDQKIDAYI 638
 DB 647 NNRSEQKHTPRDANYIFVRAHDSVQAVLANIYSKQINPKTDGFTFNDLKAQAEIYN 706
 QY 639 QDQNSTVKKYNLYNPASYAILLNKNDIPRYVYGDLYTDGGQYMEHQTRYDYDTLTLNLLK 698
 DB 707 ADIAKADKKYTYNIPAAAYATMLANKDSITRYVYGDLYTDGGQYMAEKSPYNAIDALLR 766
 QY 699 SRVYVAGGQSMOTMSVGNNNILTSVEYKGAMTATDTGDETRTQIGVYVSWNTPNLK 758
 DB 767 ARKIYAGGQDMKVTYKLG-YEIMSSVRYKGAEENOLGTAETRNQGMVLVLTANRPMK 825
 QY 759 LGVNDKVVLLHMGAAEKNOYRAAVLTITTDGVINITSQAGP---VAMTDENGDLVLSHN 815
 DB 826 LGANDRLVYVNGAAHKNQAYRPLLKSTGLATYIKDSVPAGLVRYTDNQGNTFTADD 885
 QY 816 LVVNGKEADTAVOGYANPDVSGYLAVVWVPVGSNDODARTAPSTKNSGNSAYRTNAAF 875
 DB 886 -----IAGHSTVEVSGYLAVVWVPVGSASENQDARTKASSTK-KGQXVFESSAAL 932
 QY 876 DSNVIFEAFSN---FVYPTKESERANVRIAQNADEFFASLGFTSFEMAPQYNSKDRITFL 932
 DB 933 DSQVIYEGFSNFPQPVKTPSYTNRV---IAQNAKLFKEWGITSEFAPOYVSSQDGTFL 989
 QY 933 DSTIDNGYAFTRDYLGMSEPNKYGTDEDLRNLALHAGLQVMAADVDPQIYNLPCKE 992
 DB 990 DSIENGAFEDRYDIAMSKNNKYGSLKDLMDALRALHAEISAIADWVPDQIYNLPCKE 1049
 QY 993 VATVTRDDRGNVWKNDAIINNLYVYNT-IGGGYQKYGGAFLDKLQKLYPEIFTRKQV 1051
 DB 1050 VVTASRTSYSTPRNABEYISLYAAKTRTFGNDFQKYGGAFLDELKAKYPAIFERVQI 1109
 QY 1052 STGVAIDPSQKITEKSAFYNGNTLHRSQYVTLKADG-QQYVNL-GTTTKQFLPIOLT 1108
 DB 1110 SNRKLITNEKLTOWSAKIFNGSIOGTGARYIVLDNATNTOYFYSKAGQT---FLPKDM 1166
 QY 1109 GEKKQGNFVK-GNDGNYIYFDLAGNKNVKNIFIEDSVGNMYFFDQCKMVENKHFVDVD 1167
 DB 1167 ---EITSGFRVGVDOVY--LSIGGYLAKNTFTQVGNANQWYFDKNGNMYTGEQVID-- 1219
 QY 1168 SYGKGTYFFLKNQVSPRGL-----VOTDNGTY-----YFDNYG 1202
 DB 1220 --GKK--YFFLDNLQLRHLVROGSDGHVYVYDPKQVQAFNGFYDFAGPRQDVRYPGNG 1275
 QY 1203 KWRNQITINAGAMIYITDENG 1223
 DB 1276 QMYRGLHDMYGTTFYFDEKYG 1296

RESULT 7

US-09-009-620-2

; Sequence 2, Application US/09009620A

; Patent No. 6127603

; GENERAL INFORMATION:

; APPLICANT: Nichols, Scott E.

; TITLE OF INVENTION: Substitutes for Modified Starches and

; FILE OF INVENTION: Latexes in Paper Manufacture

; FILE REFERENCE: 0357C

; CURRENT APPLICATION NUMBER: US/09/009,620A

; CURRENT FILING DATE: 1998-01-20

; EARLIER APPLICATION NUMBER: 08/485,243

; EARLIER FILING DATE: 1995-06-07

; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 2

; LENGTH: 349

; TYPE: PRT

; ORGANISM: Streptococcus mutans

US-09-009-620-2

Query Match 4.9%; Score 326.5; DB 3; Length 349;

Best Local Similarity 26.5%; Pred. No. 9.6e-16;

Matches 102; Conservative 49; Mismatches 139; Indels 95; Gaps 12;

QY 19 SAVSGFDITIKLTNDQYQALNGQLVLLRFSKAADG--NPSGDVTVDFQSKNATATGCGN 76

DB 29 TSLSGSLVKADSTDROQAVTESQASLVTTSEAAKETLTATDTSTATSQPTATVTDN 88

QY 77 PDYKVNQVNEFSGHATNQSDNDKQWIIVLVNGKEVKQLVNDTKREGAAGFNRNDVY 136

DB 89 VS-----TNGS-----TNTANTA-----NEVV 107

QY 137 KYNPAIENSSMSGFGIITLPVTVKKNVQLVHRFSNDVKTGEGNYV-----DFW 186

DB 108 KPTTSEQAKTNSDKIITTSKAVNR-----LTATGKFPANNNTAHPKTVT 154

QY 187 SELMPVKDSQKNGPLAQ-----PGLQTINGQQYVIDPTTGQPKNFLL 231

DB 155 KRIVPIPKI-----GKLQPSLSQDDIAALGNVKNIRKNGKYYY-KEDGTLOKNTAL 209

QY 232 QSGNWNIVDSGTGVTGNALEQLQFAGKVSSNE---QYRNGNAAYSDDKSTENYNGILT 288

DB 210 NINGKTFPDETGALSNNTLPSK--KGNITNDNTNSPAQYNQVYSTDVANFEHVDHILT 267

QY 289 ADTWTRPKQILKDGTTWDSKTDNRPIIMVWNPWLTOAYLYNTMKQHGULLPSALPFF 348

DB 268 AESWTRPKYILKDKGTWQSTFKQFRLPILMTWPDQETQROYVYVNAQAGTHQT-----Y 323

QY 349 NADADPAELNHYSEIVQVQNIKRIS 373

DB 324 WTATSEQLNLAAQIQTQKIEKIT 348

RESULT

US-09-268-347-49

; Sequence 49, Application US/09268347

; Patent No. 6335182

; GENERAL INFORMATION:

; APPLICANT: Loosmore, Sheena M.

; TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS

; FILE REFERENCE: 1038-860

; CURRENT APPLICATION NUMBER: US/09/268,347

; CURRENT FILING DATE: 1999-03-16

; NUMBER OF SEQ ID NOS: 54

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 49

; LENGTH: 2314

; TYPE: PRT

; ORGANISM: Haemophilus influenzae

US-09-268-347-49

Query Match

Best Local Similarity 3.7%; Score 245.5; DB 4; Length 2314;

Matches 275; Conservative 167; Mismatches 512; Indels 447; Gaps 68;

QY 18 NSAVSGFDITIKLTNDQYQALNGQLVLLRFSKAADGPNPSGDVTVDFQSKNATATGCGN 77

DB 508 NSTLNNGLTV---NNTIGGSKQIQV-----GADGKIFADVNVNVSNAKEFGTT----- 554

QY 78 DIYKVNQVNEFSGHATNQSDNDKQWIIVLVNGKEVKQLVNDTKREGAAGFNRND--Y 135

DB 555 ---RITBEIGFAD---ADGKVDKKSPLYL-----DKQLQVGGVKITKD---SGINAGDQKI 602

QY 136 YKVNPAIENSSMSGFGIITLPVTVKKNVQLVHRFSNDVKTGEGNYVDFWSELMPVKDS 195

DB 603 SNVXDATDD-----TDAVTYKQ-----LKQV 623

QY 196 FQKNGPLKQFGLQTINGQQYVIDPTTGQPKNFLLQSGNWNIVDSGTGVTGNALEL-- 253

DB 624 QQDADGALQSEFIRDEKQSEETI-----SNLYSNGVTPNTFEITFAGENGISISN 674

QY 254 QFAGKTVSSNEQYRNGNAAYSDDKSTENYNGILTADTWTRPKQIL--KDGTT----- 304

DB 675 DIARKGVK-----VGIDPINGLT-----PKLTGSDKDGKTQLVIEQ 712

QY 305 -----WDSKETDRPILMVWNPWLTOAYLYNTMKQHGULLPSALPFFNADADPAELNH 359

DB 713 VASNGTKNIIRGVSPIL-----PSITNAGGVRTTEQ--GNFITS-----DEKSKAAS 759

QY 360 YSEIVQO--NIEKRISETGNTDMLRILMHDFVTNPNMWNKDSNVNFSGIQFGGFKLYE 417

DB 760 IGDILATGTFNLKNNNSNVGVFYSTYNTV--DFIDGNATAK-----VYID 801

QY 418 NSDLTPYANSYRLLGRMPINIKQOITRGQGFLLANDIDNSNV--VQAEQLNWLILNLF 476

DB 802 ETNQTSKVYD-----VNVDEKTIE-----LTGNGKTNKIGVKTILT----- 840

QY 477 GTITANNDQANFDSVRYDADPNID--ADLMHIAQDYFNAAVGMDSDAVSNKHINILSDWN 534

DB 841 --TTNANGKATFSTDDNDALVNAKDIAENLTALKEHTTKGTADTALQT--FKVKKD-G 896

QY 535 HADPEYF--NKIGNPQLTMDPTIKNSLHGLSDATNR-----WGLDAIVHQSLADRENS 587

DB 897 ATDDETITVGKQGTQNGKTVNTLKLKENGTLVATNKGDTVTFGINTQSGLKAQGSITLN 956

QY 598 TENNVIPNYSFVRAHDNNSQDQIQNAIDVTGKDYHHTFPEDE--QKQIDA---YTDQ 641

DB 957 KDGLSIKPA-----SNEQIQ-----YCADGVKFAKVDKGNSTGTGDSRITKQ 1002

QY 642 -----NSTVKYNLYNPASVAILLTNKTIPRVYVGLYD-----GG 680

DB 1003 IGFTGANGSLDTTPHLLTKDKLVGEVEITNIGNAGKKITNTQSGDITQNSDNDATVG 1062

QY 681 QYMEHQTRYDTLNLKSRVYVA--GGQSNQZMSV----- 715

DB 1063 -----RVYDLKTE--LESKINSAAKTAQNSLHEFSVADEQGNHFTVSNPSSYDTSKTS 1114

QY 716 -----GGNNNLTLSVRYCKGAMTATDGTGDETRQGGIGVYVSNTPNLKLVNDKVLHMG 770

DB 1115 DVITFAGENGITTKVNGK-----VVRVGIDQTK--GL-----TTPKLVFG----- 1152

QY 771 AAHNQOYRAAVLTTTDDG-----VINYTSQGAPVAH-----TDEN-----GD 808

DB 1153 -----NNNGKGVIDSQDQNTITGLSNLWVND--GAGHALSQGLANDTDKTRAASIGD 1207

QY 809 LYLSSHNLVYNGKEEADTAVOGYANPDVSGGLVWVPVGSAD-----NODARTA 857

DB 1208 VLNAGFNLOGNS--EAVDFVSTYDTPVDFIDGNATATKATYTDYDTSKTVVYDVVNDKTI 1265

QY 858 PSTE-----KNSGNSAVRTNAAFDSNVLFESNFEVTPPTKESBRANVRQAQ 905

DB 1266 EWTSDKKLVGKTTTLTKTSANGNATKFSRAGQDALVKASDIATHLNTLSGDIQTAKGASQ 1325

QY 906 ADFFASLGFTSFEMAPQYNSSKDRFTFLDSTIDNGYAFTRYDRLGNSEPNKYGTDEDLRNA 965

DB 1326 ASSASAS-----YVDADGNKVIYDSTDKKYQVNDK---GOYDKNKEVAKDKL--- 1369

QY 966 IQALHKAQLOWADWVPDQIYNLPQKEVATVTRVDDRWKWKDALIINNLLVWNTIGGGE 1025

DB 1370 -----VAQAQTPD-----GTLAQMNVKSVIVKEQVNDAN----- 1398

QY 1026 YQKQYK---GAFDLKQKLYPEIFTKKQVSTGVALDSQKITESAKYFNGTNLHRGS 1081

DB 1399 --KKGIDNEDNNFINGLENAAKDKTKK---NAAVTVGOLNVAQPTPLTFAGDTGTAKKL 1453

QY 1082 GYVLKADGGQYVNLGTTTQKFLPIQLTGEKKQGNBGFVKNGDQNYFYFDLAGNMVKNFTFI 1141

Db 1454 GETTIGGQ-----TDNKL-----TDNIGVAGTDG--FTVKLAKDLT----- 1492
Qy 1142 EDSYGNWYFFDQDQKMYENK--HFVDVDSYGEKGTYPFLKNGVSFRG-----GLVQTDN 1193
Db 1493 -----NLNSVAGGTRIDEKGISFVDANGQAKANTPVLISGLDLGGKRTISNIGAAVDDN 1547
Qy 1194 GTYFDNTGKMYR-----NQINAGMI--YTLDENGKLI-----KASYNDAEY-- 1236
Db 1548 DAVNFKQNEVAKTYNNLNQNSGASLFFVVTDANGKPIINGTDCKPKQAKIGADGKYIH 1607
Qy 1237 -----PTSDVGKMLDONKL 1251
Db 1608 ANANGVPVDKDKPITDADKL 1628

RESULT 9
US-08-480-604A-10
; Sequence 10, Application US/08480604A
; Patent No. 5736139
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALLEY, BRUCE S.
; APPLICANT: PADWEE, NISHA V.
; APPLICANT: FIRCA, JOSEPH R.
; APPLICANT: STAFFORD, DOUGLAS C.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480.604A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/422,711
; FILING DATE: 14-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/405,496
; FILING DATE: 16-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REFERENCE/DOCKET NUMBER: 40,027
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2366 amino acids

; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-480-604A-10

Query Match 3.3%; Score 222; DB 1; Length 2366;
Best Local Similarity 20.4%; Pred. NO. 5.4e-07;
Matches 220; Conservative 155; Mismatches 401; Indels 304; Gaps 60;

Qy 270 NAAYSDEKSIENYNGYLTD-----WYRPQILKDGTTWTDK-----ETDMRPIL 317
Db 1459 NIPYSFYD--SEGRENGFINGSTREGLFVSELPDVLISKVYMDSPSPFGYYSNNLKDYK 1517
Qy 318 MVMWPN--TLTQAYLYNTMKQGNLLPSALPFFNADAPAEIN--HYSEIVQOQIERKISE 374
Db 1518 VTKDNVNILTGYYLK-----DDIKISLSLTQDEKTIKLSNVHLDSEGVAEILKFMNR 1571
Qy 375 TGTDLWLRTLMHDFVTNNPMWNKDSNVNFSGIFQGGFLKYENSDLTPYANSDYRLGR 434
Db 1572 KGNITSDSLM--SFL-----ESNLIKSI-----FVNELOSNIKFIIDANFTISGT 1615
Qy 435 MPINIKDQTYRGQEFLLANDIDNSPVVQAEQLNMLYLLNFGTITANND-----484
Db 1616 TSI-----GQPEFICDENDNIQ-----YFIKNTLETNYLYVGNRQNM 1656
Qy 485 -QANFDSRVYDAPDNIDADLMNTAQDYFNAAAYGMDSDAVSNKHI---NILEDWNHADPEY 540
Db 1657 VEPNYD---LDDSGDISSTVINFQKYL---YGIDS--CVNKVVISPNIYTDENITPVY 1708
Qy 541 FNKIGNPQLTMDDT--TKNSLNHGLSDATNR--WGLDAIVHQSLADRENNSTENVIPNY 596
Db 1709 ETNNTYPEVILDANYINEKINWINDLSTIRYVWSDGNDGFIKMTSEENKVSQVKI--- 1765
Qy 597 SFVRAHDNNSQDQIQONAIROVTGKDYHTFTFEDQKIDAYIQDNSTVKKYNLYNPAS 656
Db 1766 -----REVNVFKDKTLANKLSFNSDKQ-----DVPVS 1793
Qy 657 YAILTNKDTIPRYVYGD-----LYTD-----GGQYMEHQTRY--DT 692
Db 1794 EIL-----SPTSYIEDGLIGYDLGLVSLYNEKFYINFGMMVSGLIYINDSLYFKPP 1848
Qy 693 LTNLLKSRV-----KY-----VAGQSMQTSVSGNNILASVRYGKGAMTATDGTDETR 743
Db 1849 VNNLTGFTVVGDDKYYFNP INGGAA-----SIG--EILIDDKNYFQSGVLQTVGFSTE 1902
Qy 744 TQGGVYVSNP--PNLK---LGVDNKVVLHMGAAHKKQYRAAY--LFTTQGVINITS--DQ 796
Db 1903 DGFKYFAPANTLDENLEGEAIDFTGKLIIDENIYIFDNYRGAVEMKELDGEMHYFSPET 1962
Qy 797 GAPVAMTDGND--LYLSSHNLYVNGKEADTAQGVANPDVS--GYLAVVVPVGSADNQ 852
Db 1963 GKAFKGLNQIGDYKYYPNSDGVKMGKFPVSDNKHYPDSDGVKMGVYTEI-----DGK 2015
Qy 853 DARTAPSTEKNSGNSAYRTNAAFDNSVIFEAFSNFYVTPPTRESERANVRINQAADFFASL 912
Db 2016 HFYFAENGEMOIG-----VENTEDGKFFAHNEDLGN-----EGEELISYS 2057
Qy 913 GFTSEFEMAPQINSKDRFTFLDSTIDNGYAFTRDYDL-----GMSEPNKYGTDEDLRNAIQ 967
Db 2058 GILAFN-----NKIYFDDSFYAVVWGLDGLDGSKYFFDEDTAEAY-- 2098
Qy 968 ALHKAGLOVMADWVPDQIYNLPG--KEVATVTRVDRDGRNVKDA-----LINNNLYVV 1018
Db 2099 -----IGLSLIND--GOYFDDGIGIMQVGFV--INDKVFYFSDGIIESGVQINDDNYFI 2151
Qy 1019 NIIG---GGEYQKKYGGAFLLKQKLYPEIFTKKQVSTGVAIDPS--OKITEWSAKYFNG 1073
Db 2152 DONGIVQIGVEDTSDGYKYFAPANTVNDNIY-----GQAVEYSGLVRYGE--DVYFQGE 2203
Qy 1074 TNLHRSQGYVLKADGGQYYNLGTTTQKFLQIQTGEKKQKNEGFVKGNDGNYFYFVLG 1133
Db 2204 TYTITGWIYDMEINESDKYY-----FNP-----ETKACKAGINLIDDIKYYF--DEKG 2249

APPLICANT: THALLEY, BRUCE S.
APPLICANT: PADHYE, NISHA V.
APPLICANT: FIRCA, JOSEPH R.
APPLICANT: STAFFORD, DOUGLAS C.
TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
PREVENTION OF C. DIFFICILE DISEASE
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEN & CARROLL, LLP
STREET: 220 MONTGOMERY STREET, SUITE 2200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US 08/915,136
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/480,604
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/405,496
FILING DATE: 16-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329,154
FILING DATE: 25-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161,907
FILING DATE: 02-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/985,321
FILING DATE: 04-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/429,791
FILING DATE: 31-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: INGOLIA, DIANE E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPFD-01763
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 2366 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-915-136-10

Query Match 3.38; Score 222; DB 4; Length 2366;
Best Local Similarity 20.48; Pred. No. 5.4e-07;
Matches 220; Conservative 155; Mismatches 401; Indels 304; Gaps 60;

Qy 270 NAAYSYDKSIENVNGYLTD-----WYRPKQLKDGTTWTDK-----ETDMRPIL 317
Db 1459 NPTYSFVD-SEKENGFRNGSKGLFVSELPDVVLISKVYMDSKSPFGYYSNNLKDVK 1517
Qy 318 MYWPN-TLTQAYINMYKQHGNI/LPSALPPFNADAPAEIN--HYSEIVQONTKEKRISE 374
Db 1518 VITKDNVNILGYVILK-----DDIKISLSUTLODEKTIKNSVHLDSEGVAEILKFNMR 1571
Qy 375 TGNTDRLTLMDHDFVNNPMNKNKSENVNFSGIQFGGFLKYENSDLTPYANSYRLRGR 434
Db 1572 KGNNTSDSLM-SFL-----ESMNIKSI-----FVNFLQSNIKFILDANFIISGT 1615

Qy 435 MPINKDQTYRGQEFLLANDIDNSNPVVQAEQLNWLNYLLNFGTITANNO----- 484
Db 1616 TSI-----GQFEFICDENDNIQ-----YFIKENTLETNYTLYVGNRQNM 1656
Qy 485 -QANFDSVRVDAPONIDALNIAQDYFNAAAYGMDSDAVSNKHI-----NILEDNHAPDEY 540
Db 1657 VEPNYD-----LDSGDISSTVINFSQKYL---YIGDS--CVANKVVISPNYITDEINIPVY 1708
Qy 541 FNKIGNPQLTMDDT--IKNSLNHGLSDATNR--WGLDAIVHOSLADRENNSSTENVVLPNY 596
Db 1709 ETNNYPEVILVDANYINEKINNVINDLSIRYVMSDNGDNDFILMSTSEENKVSOKI--- 1765
Qy 557 SFVRAHDNNSODQIONAIRDVTGKDYHTFTFEDEQKIGIDAYIQDNSTVAKYKNLYNIPAS 656
Db 1766 -----REVNVFKDKTLANKLSFNESDKQ-----LVTD-----GGQYMEHQTRY- 1793
Qy 657 YAILLTNKDITIPRVYGD-----LVTD-----GGQYMEHQTRY- 692
Db 1794 EILL-----SFTPSYEDGLGYDLGLVSLYNEKFYNNFGMYSGLYINDSLYFKPP 1848
Qy 693 LTNLKSRV-----KY-----VAGGSMQTSVGGNNNLTSVRYGKGAMTATDGTDETR 743
Db 1849 VNNLTGFTVVGDDKYYFNPINGGAA---SIG--ETIIDDKNYFQSGVLQTCGVESTE 1902
Qy 744 TQIGGVVVSNT--PNLK---LGVNDKVVLMHGAANKQYKAAV--LTTDGVINITS-DQ 796
Db 1903 DGFYFAPANTLDENLEGEAIDFTGKLIIDENIYFDDNYRGAVEMKELDGEHMFSPET 1962
Qy 797 GAPVAMTDENG--LYLSSNHLVYNGKEADTAVOGVANPDVS--GYLAVVVPVPGASDNQ 852
Db 1963 GKAFKGLNQIGDYKYFNSDGVMOKGFVSINDKHYFDDSGVMKVGYTEI-----DGK 2015
Qy 853 DARTAPSTEKNSGNSAYRTNAAFDNSVIFAFSNFVTPAKESERANVRLAQNADFEASL 912
Db 2016 HFEAENGEMQIG-----VFNTEDGFKYFAHNEDLGN---EAGEEISYS 2057
Qy 913 GFTSFMAPQYNSSKDRTELDSTDNGYAFTRDYDL-----GMSEPNKYGTDEDLRNAIQ 967
Db 2058 GILNFN-----NKIYFDDSTAVVGMKDLSDGSKYFDEDETAEAY- 2098
Qy 968 ALHKAGLQVMADWVPDQIYNLPG-KEVATVTRVDRGNVWKDA-----IINNLYVV 1018
Db 2099 ----IGLSLND--GQYFENDGIMQVGFV--INDKYFYSDSGILIESGVQNDIDNIFYI 2151
Qy 1019 NTIG--GGEYQKKGAGFLDKLQKYPELFTKKQVSTGVAIDPS--QKITESSAKFYNG 1073
Db 2152 DDNGIVQIGVEDTSDGYKFEAPANTVMDNIV-----GQAVEYSGLVRYGE-DVYFGE 2203
Qy 1074 TNLHRSYVVLKADGGYVNLGTYTKQFLPIOLTGEKKQNGEFGVKGNDGNTYFYDLAG 1133
Db 2204 TYTETGWIIDMENESDKYY-----FNP-----ETKACKGINLIDDIKYF-DEK 2249
Qy 1134 NMVKNTFIEDSVGNWYFFDQDGKM-----VENKHFDVDSYGEKGYFFLKNVSGFRGG 1187
Db 2250 --IMRTGLISFENNYYFENGEMQFCYINIEKMFY---FGEDGV-----MQIG 2294
Qy 1188 LVQTDNCTYTFEDNYGKMYRN---QTINAGMIYT---LDENGLKIKASYNDAEYPTST 1240
Db 2295 VFNTPDGFKYFAHONTLDENFEGESIN---YTGWLDLDE-----KRYFTD-EYIAAT 2343

RESULT 12

US-08-480-604A-6
Sequence 6, Application US/08480604A
Patent No. 5736139
GENERAL INFORMATION:
APPLICANT: KINK, JOHN A.
APPLICANT: THALLEY, BRUCE S.
APPLICANT: PADHYE, NISHA V.
APPLICANT: FIRCA, JOSEPH R.
APPLICANT: STAFFORD, DOUGLAS C.
TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
PREVENTION OF C. DIFFICILE DISEASE

Qy	1189	VQZDNGTYPDNKGKVRN--QTINAGAMIYFLDENGKLKASYNSDA-----	1234
		: : : : : : : : : : : : : : : :	
Db	2484	-QTINKKYFNTSIASTGYTIISGRHFY-----FNTDGIQIGVFKGPDG	2530
Qy	1235	-EY--FTSTDVGKMLDQ 1248	
		: : : : : : : : : : : : : : : :	
Db	2531	FEVFPANTDANNIEGO 2547	

```

RESULT# 13
US-08-405-496A-6
: Sequence 6, Application US/08405496A
: Patent No. 5919665
: GENERAL INFORMATION:
: APPLICANT: WILLIAMS, JAMES A.
: TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM
: TITLE OF INVENTION: NEUROTOXIN
: NUMBER OF SEQUENCES: 30
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: MEDLEN & CARROLL, LLP
: STREET: 220 MONTGOMERY STREET, SUITE 2700
: CITY: SAN FRANCISCO
: STATE: CALIFORNIA
: COUNTRY: USA
: ZIP: 94104
:

```

```
,  
,  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentID Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/405.496A  
FILING DATE: 16-MAR-1995  
,
```

CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329,154
FILING DATE: 25-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161,907
FILING DATE: 02-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/985,321
FILING DATE: 04-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/429,791
FILING DATE: 31-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: INGOLIA, DIANE E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPBD-0134
TELECOMMUNICATION INFORMATION:

```

/ TELEPHONE: (415) 705-8410
/
/ TELEFAX: (415) 397-8338
/
/ INFORMATION FOR SEQ ID NO: 6:
/
/ SEQUENCE CHARACTERISTICS:
/
/ LENGTH: 2710 amino acids
/
/ TYPE: amino acid
/
/ TOPOLOGY: linear
/
/ MOLECULE TYPE: protein
/
/ US-08-405-496A-6

```

Query Match 3.18; Score 209; DB 2; Length 2710;
Best Local Similarity 18.1%; Pred. No. 5.7e-06;
Matches 263; Conservative 198; Mismatches 52; Indels 484; Gaps 73;

```

QY 60 NTVTDQFSKNIATTTGNFDYKVNQGVESGW-ATMSNDKDSQMT-----IVL 109
      | : : : | : : | : : | : : | : : | : : | : : | : : | : : |
Db 1307 NEIRNLKSYSDGAGGTYSLL-----LSEYPTSTWNLNKDGLWFNIDNREIRSI 1358
      | : : : | : : | : : | : : | : : | : : | : : | : : | : : |
QY 110 VNGKEVKRLVNDTTEGAAGFNRRNDYKVNPAIENSSMSGFG-----ITLPTV 160
      || : : | : | : | : : | : : | : : | : : | : : | : : |

```

Db	1359	ENGITKGGKLIKOV-----LSKIDINKNKLLIGNOTID-FSGDIDNKRIFLTCELD	1411
Qy	161	KNE-- --NQVLVHRESDVKTEGNYVDFWSELMPVKDSFKQGNGLPKQKGLQTINGQOYY	217
Db	1412	KISLIIEINLVAKSYSELLSGDANYL-----ISNLSNTEK-----INTLGLDSKNIAVY	1462
Qy	218	IDPTGQPRANFLLOSNNWIIYFSDTGVGTNALEOFAGKTVSSNEQYRNGNAAYSDD	277
Db	1463	TD-----ESNNK--YFGAISKTQSAKI-IHYKKDSKNITLEFYNDSTLHFNKSD	1507
Qy	278	KSIENWNGYLTADTWYRPKQIADGTTWDSKETDMRPLIWWNTLTQAYYL-----	331
Db	1508	FIABDIINFWKDDI-----NTITGKYYVDNNTDK	1536
Qy	332	-----NYMKQH-----NLLPSAIPTF-NADADPAELNHS-----	361
Db	1537	SIDFSISLVSKNQVKVNGLYINESVYSYLDVFKNSDGHHTNFEWNLFLDNISFWKLF	1596
Qy	362	EIVQOIKERISGETNYT-----DMLRTLMHDFVTN-----	399
Db	1597	FENINFYIDKTYFLVGTNIGYVEFICDNNKNIDIVFGWKTSSSKSTIFSGNGRNWVE	1656
Qy	393	PMWKO-----SENWFS-----GIOFGGFLKYENSOL-----TPYANSDY	429
Db	1657	PIYNPDTGEDISTLSDFSZEPLGYD-----RYTNKVLIAPLDLYTSLINTNTYISNEY	1710
Qy	430	-----RLGCRMPINIKDQTY-----RGOEFLANDIDNSN-PVQAOELMWLY	472
Db	1711	YPEIIVLPNTFHKKNVINLDSSFEYKHWSTGSDFLVRYLEESNKKILQIKRIKI--	1768
Qy	473	LLNFGTITANNQANFDSVRVDAPDNIDADLMIADQY--FNAAVGMDS--AVSNK	525
Db	1769	-----LSNTQ-SFNKASIDFKDKKLSLGYIMSNKSPNSENELDRDHGLGFIIDNK	1819
Qy	526	H-----INILEDWNHADPEYFNKIGNQOLTUMDDTIKNSLHGL-----	563
Db	1820	TYYYDEDSKVLGVLINNSLFFEDPIELFNLVGTWOTINGKYYFIDNTGAALSYKIIN	1879
Qy	564	-----SDATNRWGL-----DAIVHQSLADRENSTENVIPNYSFVRAHDNNSDQION	612
Db	1880	GKHFYFNNDGVQMQLGVFKPGDGEFYEPAPANTQNNIEGQAI-----VYQS	1924
Qy	613	AIRDVTGQYHTTFFDEDKGI-----DAYIQDQNSTVKKYNLYNIPASAYAILLTK	664
Db	1925	KFLTLANGKYY--FONNSKAVTGWRIINNEKYYFNPNNAIAAVGL-----QVIDNNK	1974
Qy	665	DTIPRVYIGB---LYVTDGQYMEHCTRYDYDLTNLLSKRVKYVAGGOSWQTMVSQGN	720
Db	1975	-----YFNPDTAILISKGWQTVNGSRYYFDDTATAFNGYK-----ITDGKHF	2017
Qy	721	ILTSVRYGKGAMTADTG-----TDETRQQIGVWVS-----NTPNLKLGVDNKVV	766
Db	2018	YFSDCVVKIGVSTSGNGFEYAPANTYNNIEGQAIYVQSKFELTNCKKYFFONNSKAV	2077
Qy	767	LHGAHKNQOY-----RAAVLTGTDGVINVTSDQGPVA---MTDENGDLYSLSHN--	815
Db	2078	TGLQITDSKYYFENTNABAAATGWOTIDGKYYFENTNABAAATGWOTIDGKYYFENTN	2137
Qy	816	-----LVVNGKEEADTAQGVANPDVSGYLAWVPVGSADQDARTAPSTEKKSNGSAY	869
Db	2138	IASGTYYTLNGK-----HFYFNTD-GIMOIGVFKG--PNGFEYFAPA--NTDANNI	2183
Qy	870	RTNAAFDSNVIFEAFSNFYVTFKESERANVRIAQADFF-----ASLGFTSFEWAPQ	922
Db	2184	EQQAILYQNEFLTNGKYYFGSDSKAVTGWRIINNKYYFNPNNAIAAHLCTI-----	2238
Qy	923	YNSKQRTFLDSTIDNGYAFTRDGLGMSPEPNKYGTDEDLRNAIQALHKA--GLQVMADW	980
Db	2239	-NNDKYYFSYDGLIQNGYITIER-----NNFYFDANNESKHWYTVFKPGNGFEYFA--	2288
Qy	981	VPDQIYNLPGKEVATVTRVDRGRNWKDAIINNLYVNTTIGGGEYOKKY-----	1030
Db	2289	-PANTN-----NNEGQAIYVQNKFL--TLNG--KKYFYDMSKAVT	2326

QY 1031 GGAFDQKQKLPYELFTKQKQVSGVAID-----PSQKITKESA-----KYPNGTUIL 1077
Db 2327 GWQTDG-KKYIFNLTAETAGWQTDIDGKKYYFNLTAETAGWQTDIDGKKYYFNTNTE 2385
QY 1078 HRCSGVLKADGQVYINLCTTTPKQLPTOLTCEKQKQNGEFGVKGNDGNVYF-----YO 1130
Db 2386 IASTGVT-SINGKHEY-----FNTDGMQI-----GYFKGPGGEYFAPANTDANN 2430
QY 1131 LAGNV--KNTFDSVGNWFFDQDGRMYENKHFVDVDSYGEKGTFFLANGVSFRGGL 1188
Db 2431 IEGQAILQNKFLTN-GKKYFSGDSKAVTGLRTIDCKY-----YFNTNTAVAVTGW- 2483
QY 1189 VOTDNGTYFEDYNGRWNRN--QTINAGAMIYTLDENGKILIKASYNDA----- 1234
Db 2484 -OTINGKKYYFNTWNTSIAGTYITISGKHEY-----FNTDGMQIGVFKGPDG 2530
QY 1235 -EY--PTSDVKGMLDQ 1248
Db 2531 FEYFAPANTDANNISQ 2547

RESULT 14
US-08-915-136-6
; Sequence 6, Application US/08915136
; Patent No. 6290960
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALLEY, BRUCE S.
; APPLICANT: PADHYE, NISHA V.
; APPLICANT: FIRCA, JOSSEPH R.
; APPLICANT: STAFFORD, DOUGLAS C.
; TITLE OF INVENTION: VACCINE, DOUGLAS C.
; TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESS: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/915,136
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/480,604
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/405,496
; FILING DATE: 16-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-01763

TELECOMMUNICATION INFORMATION:

; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2710 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-915-136-6

Query Match 3.18; Score 209; DB 4; Length 2710;
Best Local Similarity 18.1%; Pred. No. 5.7e-06;
Matches 263; Conservative 198; Mismatches 512; Indels 484; Gaps 73;

QY 60 MTVDQPSKNTATGGNFYVYKVGNGQVFEFSGWH-ANMQSNDKDSQIT-----IIVLPVTV 160
Db 1307 NEIRNKLSYFSDGAGTYSL-----LSSYFISTNINLSKDDLMIFNIDNEVREISI 1358
QY 110 VNGKEVAKQLVNDIKEGAAGFNRRNDVYKVPAINSSMSGFQ-----ITLPVTV 160
Db 1359 ENGATKKGKLIKDV-----LSKIDINKKLIIGNQYTD-FSGDINKDRYIFLTCELD 1411
QY 161 KNE---NVQLVHRESNDVKUGENYVDFWSELMPVKDSFOKNGPLQFGLOTINGOOXY 217
Db 1412 KISIIIEINLVAKSYLLSGDKNYL-----ISNLSNTIEK-----INTLGDSKNIAINY 1462
QY 218 IDPTTGPRKNFLQSGNNWYFDSGTGVTGNALELOFAKGTVSSNEQYNGNAAYSYDD 277
Db 1463 TD-----ESNNK--YFGAISKTSQKSI-IHYKDSKNILEFYNDSTLEFNSKD 1507
QY 278 KSIENVNGYLTADTWYRPKQILKDGTTWDSKETDRPILMVPMTLTQAYYL----- 331
Db 1408 FIAEDINVFEMKDDI-----NITKYYVNDNTDK 1536
QY 332 -----NYMKQH-----NLPSALPFF-NADADPAELNHY- 361
Db 1537 SIDFSISLVSKNQVKNGLYLNESVYSYLDFAVNSDGHHTSNFMNLFNDNISFWKLF 1596
QY 362 -EIVQONIEKRISGTNT-----DMLRTLMHDFVTN-----N 392
Db 1597 FENINFDKRYFTLVGNTLNGYVEFICDNKNKNDIYGEWKISSKSTIFSGNGRNVVE 1656
QY 393 PMWKKD-----SENVNFS-----GIQFGGFLKYENSDL-----TPYANSDY 429
Db 1657 PIYNPTGEDISSLDFSYEPLYGID-----RYINKVLIAPDLYTSLININTNYSNEY 1710
QY 430 -----RLGMRPINIKDQTY-----RQGEFLANDIDNSN-PVVOABOLWLYY 472
Db 1711 YPELIVLNPTWTFHKKNVINLDSSEFYKWNSTEGSDFILVRYLEESNKKILQKIRKI-- 1768
QY 473 LLNFGTITANDQANFDSVRVADPNIDADLMNIAQDY--FNAAAGMDSD-----AVSNK 525
Db 1769 -----LSNTQ--SEFNKMSIDFKDKLKSIGYIMSEKSFENSENELDROHLGFKITDNK 1819
QY 526 H-----INILEDWNHADPEYFNKIGNPQITMDOTIKNSLNHGL----- 563
Db 1820 TYYDEDSKLVKGLININNSLFPDTEFNVLVGTWQTINGKYYFDINTGAALTSYKIIN 1879
QY 564 -----SDATNRWGL-----DAIVHQSLADRENNSNVVPIPNYSFVRADNNSQOIQW 612
Db 1880 GKHFYFNDGVMLGVFKGPDGGEIFAPANTQNNIEGQAI-----VYQS 1924
QY 613 AIRDVTGKDYHTFTFEDEQKI-----DAYIQONSTVKYKYNLYNIPASVAILLTK 664
Db 1925 KFLTLNGKKYY-----FONNSKAVTGWRIINNEKYYFNPNNIAAVGL-----QVIDNNK 1974
QY 665 DTPRVYIGD-----LYTDGQYMEHQTRYDTLTNLKSRKYVAGSQMOTMSVGGNNN 720
Db 1975 -----YFNPDTAILISKWQTVNGSRYYFDTTATAFNQYK-----TIDGKHF 2017
QY 721 ILIUSVRYGKGAWTATDGS-----TDEIRWQIGIGVVVS-----NTPNLKLGVDNKVV 766

Db 2018 YPDSVVKIGVFSTNGFEYFAPANTYNNNIEGQAIYVOSKFLNLNGKYYFEDNNSKAV 2077
Qy 767 LHMGAHKNQY-----RAVLTTTCVINYTSQCAPVA---MTDENGOLYSSHN- 815
Db 2078 TGLQTDISKYYFNTAATAGMTIDGKYYFNTAFAATGQOTIDGKYYFNTA 2137
Qy 816 -----LVVNGKEADTAVQGYANDVSGYLAVWVPVVGASDNODARTAPTEKNSGSAY 869
Db 2138 IASTGYTIINGK-----HFVFNVD--GIQIGYFKG--PNGFPYFAPA---NFDANNI 2183
Qy 870 RTNAAFDSNVIPFARSNVYPTKESERANVRIRIQAADFF-----ASLGFTSFEMAPQ 922
Db 2184 EQQAIIYQNEFTLNGKYYFGSDSKAVTGWRIINNKYYFNPNNIAIAIHLCTI----- 2238
Qy 923 YNSKDRFTLSDTIDNGYAFTRDYDLGHSSEPKYGTDEDLNAQALHKA--GIQVMDW 980
Db 2239 -NDKYYFSYDILQNGYITIER-----NNFYDANNESKAVTGVFKGPNGEYFA-- 2288
Qy 981 VPDQIYNLPKGEVATVTRVDDRGVWMDAIINNLLYVYVNTIGGGEYQKKY----- 1030
Db 2289 -PANTHN-----NNIEGQAIYVQNKFL--TLNG---KKYFEDNSKAVT 2326
Qy 1031 GGAFLDKLQKLYPELFTKKQVSTGVAID-----PSQKITEWSA---KYFNGTHIL 1077
Db 2327 GQOTIDG-KKYYFNLTAEAAATGQOTIDGKYYFNLTAEAAATGQOTIDGKYYFNTF 2385
Qy 1078 HRGSYVLKADGGQYVNLGTTKQPLQLAGKQKQNGEFGKNDGNYYF-----YD 1130
Db 2386 IASTGYT-SINGKHEY---FNWDGIMQI-----GVFKGPNGEYFAPANTDANN 2430
Qy 1131 LAGNVV--KNTFIEDSVGNWYFFDQGRKVENKHFFVDVDSYGERGTFYFLKNGYSFRGGL 1188
Db 2431 IEGQAIYQNKFLTLN-GKKYYFGSDSKAVTGLRTIDGKKY-----YFNTAVAVTGW- 2483
Qy 1189 VQTONGTYFFDNYGHWVN--QTNAGAMITLDENGKLIKASNSDA----- 1234
Db 2484 -QTINGKYYFNTNTSTASTGYTISGKHEY-----FNTDGIQIGVFKGPDG 2530
Qy 1235 -EY--PTSTDVGMKMDQ 1248
Db 2531 FEYFAPANTDANNIBGQ 2547

RESULT 15

us-09-377-155-33
; Sequence 33, Application US/09377155
; Patent No. 6197312
; GENERAL INFORMATION:
; APPLICANT: PEAK, Ian Richard Anselm
; APPLICANT: JENNINGS, Michael Paul
; APPLICANT: MOXON, E. Richard
; TITLE OF INVENTION: NOVEL SURFACE ANTIGEN
; FILE REFERENCE: 065064/0128
; CURRENT APPLICATION NUMBER: US/09/377,155
; CURRENT FILING DATE: 1999-08-19
; PRIOR APPLICATION NUMBER: PCT/AU98/01031
; PRIOR FILING DATE: 1998-12-14
; PRIOR APPLICATION NUMBER: GB 9726398.2
; PRIOR FILING DATE: 1997-12-12
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 33
; LENGTH: 2353
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
us-09-377-155-33

Query Match 3.1%; Score 208; Db 4; Length 2353;
Best Local Similarity 19.7%; Pred. No. 5.4e-06;
Matches 282; Conservative 156; Mismatches 447; Indels 548; Gaps 74;

Qy 29 KLTNDQ--YQALNGQLQVLLRFSKAADGNPSGD-----NTVTDQFSNNYATQGNFD 78
Db 1088 KLVNABGLATALN-NLSWTAKADYADGESGETDQEVKAGDKVTFKAGN----- 1137
Qy 79 YVKVNGNVEF-----SGNHATNQSDKDSQMIIVLVNGKEVKRQ 118
Db 1138 -LKVKQSEKDFYSIQDTLGLTSLITLGGTANGNDGTGVINKDG-LTITLANGAAAGTD 1195
Qy 119 LVND-----TKEGAAGFNRN-----DVKY-----VNPATENSMSFQG 152
Db 1196 ASNGNTISVTVDGTSAGNKEITNVKSAIKYKOTQNTADETODKEPFAAIVKNAVEFVG 1255
Qy 153 IITLPVYKNNVQVHVFHSNDV---KTGEGNYVDVWSELMPVKDSFKQKNGPLKQGLQ 209
Db 1256 KNGATVSAKTND-NGKKTVTIDVAEAKVGDGLEKD-----TDGKIKLKVDN 1300
Qy 210 TINGQQYIIDTTPQPRKFNLLQSGNNIYFDSDTGVGTNALELQPAKGTVSSNEQVRNG 269
Db 1301 TDGNNLLTVDAIKGAS---VAKGEFNAVITDATTAQGTNANE---RGKVVYVKG--NG 1350
Qy 270 NAAYSDDKSTENYVNGYLTAQTYWRPKQILKDGTTWDSKETDMRPILAVWNPNTLQAY 329
Db 1351 ATATETDKKKVATYGDVAKA-----INDAATFVKVENDDSATI----- 1388
Qy 330 YLNTMKQHGNLPSALPFENADADPAELNHYSEIVQONIEKRISSETGNTDHLRTLMHDFV 389
Db 1389 -----DDSP-----TDDGANDALKA---GDTL 1407
Qy 390 T-----NHPMMNKDSYENFSGIOFGGFLKYENSDLTPYANSDYRLGLRMPINIKOQTYR 445
Db 1408 TLKAGKNLKVROCKNITFA----- 1427
Qy 446 QOEFLANDIDNSPVVQAQLWLYLLNFGTITANDQANFDSVRVADPNIDADIMN 505
Db 1428 -----LANDLSVKSATYSK-----LSLGT---NGKNVNTS-----DTKGLN 1462
Qy 506 IAQDYFNAAYGMSDAYSNNKHINLEDMNHADPEYFNKIGNPQITMDDTIKNS-----L 559
Db 1463 FAKD---SKTGDD---ANTHLNGIAS-----TLDTLLNSGATTNLG 1498
Qy 560 NHGLSDATNR-----WGLDAIHQSLADRENNSTENVIPNYSFVRAHDNNSQ 607
Db 1499 GNGITDNEKKRAASVKDVLNAGMNVGVKPAAS---ANNQVENI---DFVATYD--- 1545
Qy 608 DQIQNAIROVTG-KDYHTFTFE--DEQKIGIDAYIQDQNSVTKKYNLYNIPASYAILLTNK 664
Db 1546 -----TVDFVSGDKDFTTSVTVESKONGKRTSVIKIGAKTSVIKDHN-----GKLTGK 1592
Qy 665 DTIPRVYIGDLYTGGQYMEHQTRYDITLNLKSKYKVA---GGQSMOTMSVGGNNNI 721
Db 1593 E-----LKDANNNGVTVE--TDGKDBGNGLVTAKAVIDAVNKAQWRVKTGTGANGQND 1644
Qy 722 LTVSVYKKGAMTADTGT-----DETFTQIGVVVSVNTPNLKLGVDNDKVVYVLMHGAH 773
Db 1645 FATVASGTNYTTFADGNGTTAEVTKANDGSI TVKYNVKVD--GLKLD-GDKIVAD----- 1696
Qy 774 KNOQYRAAVLTITTDGVIINTYSDQAPAMTDENGDLVLSHNLVYVNGKEADTAVOGYAN 833
Db 1697 -----TTVLTVADGKVT-----AP-----NNGD-----GKKFVDAS----- 1722
Qy 834 PDVSGYLAVWVVPVVGASD--NODARTAPSTEKNSG-----NSAVRTNAAPDSNVIFPASN 886
Db 1723 -----GLADALNKLWATATAGKEGTGEVDPPANSACQEVKAGD-KVTFKAGD- 1767
Qy 887 FVTPTKESERANVRIAQNA-DFEASL-----GFTSFEMAPQYNSSKRTFLDST---IDN 938
Db 1768 -----NLKIKQSGKDFYSILKELKDLTSVE---PKDANGCGTSESTKITD 1811
Qy 939 GYAFTDRYDLGKSEPNKYGTDEDLRNAIQALHKA-----GLOVMADWVPDQIYNLPGKE 992
Db 1812 GLTITPANGAAGAAGANTANTISVTKDCISAGNKAVTNVVSGLKFKGDOG-----HTLANGT 1866
Qy 993 VA-----TVTRVDRGNWVKDAIINNLLYVYVNTIGG-----GEYQKKY 1030

Db 1967 VAFPEKHYDNYKDLNLEKGDNDNPVADNTAATVGLRGWISADKTTGERNQEX 1926
QY 1031 GCAFLDKLKLPEITFKQVSTGVALDPSQKITEKSAKYNFTNLL--HRSQGVYK-- 1086
Db 1927 MAQVRNANE-----VKFKSGNGINV-----SGKTLNGTRVITPELAKGEVKS 1970
QY 1087 -----ADGGO-----YYN-----LGTTKOFLP-----IQLTGEKKQ-- 1113
Db 1971 EFTVKNADGETNLYKVGDMYKSKEDIDPATSPWTKTEKYKVENGVKSANGSKTEVT 2030
QY 1114 ---GNEGFVKGNDGNYFYVDLAGNMYKNTF---IESVGNWTFDDGKAVP----- 1159
Db 2031 LTNKSGGYVTCN-----QVADAIAKSGFELGLADAAEAKFAESAKDKQLSKDKAET 2083
QY 1160 -NKH-----FVDVSYGKGYFFLKNQVSF-----RGGLVQ 1190
Db 2084 VNAHDKVRFANGLNKKVSAATVSTGANGDKVTTTFVKYDVELPLTQIYNTDANGNKVIK 2143
QY 1191 TDNGTYFYDMYGRMVRNQTINAGAMITYLDENG-KLIKASYN-SDAEYPTSTD 1241
Db 2144 KADSKWYELNADGTASNKEVTLG-----NVDANGKVKVKTENGADKWIYTNAD 2192

RESULT 16

US-08-913-942-4
; Sequence 4, Application US/08913942
; Patent No. 6200578
; GENERAL INFORMATION:
; APPLICANT: St. Gene, Joseph
; APPLICANT: Barenkamp, Stephen J.
; TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr Hohnbach Test Albritton & Herbert LLP
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/08/913,942
; APPLICATION NUMBER: US/08/913,942
; FILING DATE: 29-DEC-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/409,995
; FILING DATE: 24-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PC/US96/4031
; FILING DATE: 22-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Vance, Dolly A.
; REGISTRATION NUMBER: 39,054
; REFERENCE/SOCKET NUMBER: 8-61053-1/RFT/RMS/DAV
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2353 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-913-942-4

Query Match

3.1%; Score 208; DB 4; Length 2353;

Best Local Similarity 19.7%; Pred. No. 5.4e-06;
Matches 282; Conservative 156; Mismatches 447; Indels 548; Gaps 74;
QY 29 KLTNDQ--YQALNQLQVLLRFRSKAADGNPSGD-----NTVTDQFSKNYATTGCGND 78
Db 1088 KLVNAEGLATALN- NLSWTAKADKIADGSEGETDQEVKAGDKVTFKAGN----- 1137
QY 79 YVYNGNGQVEF-----SGNHATNQSNDSDQWIIIVLVNGKEVKRQ 118
Db 1138 -LKVKQSEKDTYSIQDTLGTGLTSLTGLGTANGENDTGTIVKDG-LTITLANGAAAGTD 1195
QY 119 LVND-----TKEGAAGERN-----DYK-----VNPALSENSMSGFG 152
Db 1196 ASNGNTISVTKDGISAGNKEITNYKSALKTYKOTQNTADETQKFEHAAVKANAEYFVG 1255
QY 153 IITLPVTVKNEENVQLVHRFSNDV---KTGECNVDFENSELMPVKQFKGNGPLKQGLQ 209
Db 1256 KNGATVSKNTDN-NCKHTVTFIDVAEAKVGGLKED-----TDGAIKLKVDN 1300
QY 210 TINGQQYVIDPTTGPQRKFNLLQSGNNWYFDSGTGVTNALLEQFAKGTVSSNEQYRNG 269
Db 1301 TDGNLLTVDATKAS-----VAKGEFNAVITDAFTAQGTANAE-----RGKVVVYKGS--NG 1350
QY 270 NAAYSYDDKSIENYNGYLTADTWYRPQILKDGTTWDSKETMRPILMWVWENTLTQAY 329
Db 1351 ATATFTDKKVVATVDVAKA-----INDAATPVKVENDDSATI----- 1388
QY 330 YLNTYMKHGNLLPSALPFFFNADADPAELNHYSEITVOQNIKRISETGNTDMRLTMDHVF 389
Db 1389 -----DDSP-----TDDGANDALKA--GDTL 1407
QY 390 T-----NNPMWNKDSNVNFSGIQFGGFLKYENSDDLPPYANSDYRLGLRMPINKDQTYR 445
Db 1408 TLKAGKLNKVRHDKNITFA----- 1427
QY 446 GQEFLLANDIDNSNPVQAEQLNMLYLLNFGTITANNQANFDSVRVADPNIDALMN 505
Db 1428 -----LANDLSVKSATVSDK-----LSLGT-----NGKNVNTS-----DTKGLN 1462
QY 506 IAQDYFNAAYGHDSAVSNKHINLLEDAHNDHPYFNKIGNPQLTMDDTIKNS-----L 559
Db 1463 FAKD-----SKTGDD-----ANIHLNGIAS-----TLTDTLLNSGATNLG 1498
QY 560 NHEGLSDATNR-----WGLDAIVHOSLADRENNSTENVIPNYSFVRADNNSQ 607
Db 1499 GNGITDNEKKRAASVKDVLNAGNVRGVKPPAS-----ANNOYENI-----DFVATYD-- 1545
QY 608 DOIQNAIRDVTC-KDYHHTFE--DEQKIDAIYQDQNSFVKKYNLNIPIASAYAILTNK 664
Db 1546 -----TVDFVSGDKDTTSVTVESKDKGRTYKIGAKTSVIKDHN-----GKLFYTK 1592
QY 665 DTIPRVYTGDLATDGGQYMEHOTRYDTLNLKSRVKYVA---GGOSMOTMSVGGNNNI 721
Db 1593 E-----LKDANNNGVYVTE--TDGKDSGNGLVTKAVIDAVNKGHRVKTGTGANGQDD 1644
QY 722 LPSYRYGKGMATPTDGT-----DETRTQGGIGVVVVSNTPNLKLGVNDKVVLMHGAH 773
Db 1645 FATVASGTVNTEADGNGTTAEVTKANDGSITVKYKVAAD--GLKLD--GOKIVAD----- 1696
QY 774 KNOQYRAVLTITTDGVINVTSDQGPAPVAMTDENGDIYLSHNLVYNGKEEADTAVQCYAN 833
Db 1697 -----TTVLTVADGKVT-----AP-----NNGD-----GKKFVDAS----- 1722
QY 834 PVSYGLAVWVPVGASD--NODARTAPSTEKNSG-----NSAYRTNAAFDSNVIFEAFSN 886
Db 1723 -----GLADALNKL-SWTATAGKSGTGBVDFANSAGQEVKAGD-KVTFKAGD- 1767
QY 887 FVYTPKESERANVRIAQA-DEFFASL-----GFTSFEMAPQYNSKDRDTLDTST--IDN 938
Db 1768 -----NLAKIKQSGKDTYTSLSLAKELDLTSVE-----FKDANGTSGESTKITKD 1811
QY 939 GYAFTDRYDLGMSNPKNYCTDDELRNIAQLHKA-----GLOVWADVVPQIYINLPCKE 992

Db 1812 GLTITPANGAGAAGANTANTISVTNDGDISAGNKAVTNVSGLKKEFGDG-----HTLANGT 1866
QY 993 VA-----TVTRVDRGNWKKDAIINNLLVVTWIGG-----GEYQKKY 1030
Db 1867 VADFKKHYNAYKDLTNLDEKAGNNPTVADNTAATVAGDLRGLGHWISADKTGTGEPNOEY 1926
QY 1031 GGAFLDKLQKLYPEFTKOVSTGVAIDPSQKITBMSAKYFNGTNIL--HRGSGYVLK-- 1086
Db 1927 NAQVRNANE-----VAFKSGNGINV-----SGKTLNGTRVITTEFELAKGEVVKSN 1970
QY 1087 -----ADGGQ-----YYN-----LGTTTKQFLP-----IQLTGEKKO-- 1113
Db 1971 EFTVKNADGSEINLVKGDYMSKEDIDPATSKPMTGATEKXKVENGVKSANGSKTEVT 2030
QY 1114 ---GNEGPFVGNDCYNYFDLAGNMVKTFF---IBDSVGNWYFFDQDGKMYE----- 1159
Db 2031 LTNKSGYVTGN-----QVADAIKSGFELGLADAAEAKAFSAESAKQLSKDKAET 2083
QY 1160 -NKH-----FVDVDSYGEKTYEFLKNGYSE-----RGGVLQ 1190
Db 2084 VNAHDKVRFANGLATKVSAATVESTDANGDKVTTTFVKTDELPUTQIYNTDANGNKIVK 2143
QY 1191 TDNGYTYFDNCKMVRNKTINAGAMIYTLDBNG-KLIKASYN-SDAEYPTSTD 1241
Db 2144 KADGWYELNADGTASNKEVTILG-----NVDANGKVKVKTENGADKWTYNAD 2192

RESULT 17

US-09-669-974-33

; Sequence 33, Application US/09669974

; Patent No. 6333173

; GENERAL INFORMATION:

; APPLICANT: PEAK, Ian Richard Anselm

; APPLICANT: JENNINGS, Michael Paul

; APPLICANT: MOXON, E. Richard

; TITLE OF INVENTION: NOVEL SURFACE ANTIGEN

; FILE REFERENCE: 065064/0128

; CURRENT APPLICATION NUMBER: US/09/669,974

; CURRENT FILING DATE: 2000-09-26

; PRIOR APPLICATION NUMBER: US 09/377,155

; PRIOR FILING DATE: 1999-08-19

; PRIOR APPLICATION NUMBER: PCT/AU98/01031

; PRIOR FILING DATE: 1998-12-14

; PRIOR APPLICATION NUMBER: GB 9726398.2

; PRIOR FILING DATE: 1997-12-12

; NUMBER OF SEQ ID NOS: 33

; SOFTWARE: patentIn Ver. 2.0

; SEQ ID NO 33

; LENGTH: 2353

; TYPE: PRT

; ORGANISM: Haemophilus influenzae

US-09-669-974-33

Query Match 3.18; Score 208; DB 4; Length 2353;

Best Local Similarity 19.78; Pred. No. 5.4e-06;

Matches 282; Conservative 156; Mismatches 447; Indels 548; Gaps 74;

QY 29 KLTNDQ--YQALNGLOQLVLLRFSAADGNPSGD-----NTVTDQSKNYATGGNFD 78
Db 1088 KLVNAGGLATALN-ILSWTAKADYADGESBGETDQEVKAGDKVTFKAGKN----- 1137
QY 79 YKVGNGQVEF-----SGWHATNQSNQDKDSQNTIIVLVNGKEVKRQ 118
Db 1138 -LKVNQSEKDFYSIQDLTGLTSITLGGTANGRNDGTTVINKDG-LTITLANGAACTD 1195
QY 119 LVND-----TVEGAGFNRN-----DVYK-----VNPAINENSWSGFQG 152
Db 1196 ASNGNTISVTKDGISAGNKEITNVKSAKLTQKDTQNTADENQDKFEHAAVKNANEVEFVG 1255
QY 153 IITLPTVTKNENVLVHFRFSNDV---KTGEGNYVDWFNSLMPVKDSQKNGPLKQFGLQ 209
Db 1256 KNGATVSARTON-NGKHTVTTIDVAEKYVDGLBKD-----TDGKIKLKYDN 1300

QY 1160 -NKH-----FVDDSYGEKTYFLKNGVSF-----RGGLVQ 1190
 Db 2084 VNAHDKVRFANGUNTKVSAATVESTDANGDKVTTTFVKYTDVDELPTQYINTDANGNKIVK 2143
 QY 1191 TDGTYTYFONYGKVRNQINAGAMITYILDENG-KLIKASYN-SDAETPTSTD 1241
 Db 2144 KADGKWFELNADGTASKEVTLG-----NVDANGKVKVKTENGADKWTYTNAD 2192

 RESULT 18
 US-08-470-260-3
 ; Sequence 3, Application US/08470260
 ; Patent No. 6077706
 ; GENERAL INFORMATION:
 ; APPLICANT: Covacci, Antonello
 ; APPLICANT: Bugnoli, Massimo
 ; APPLICANT: Telford, John
 ; APPLICANT: Macchia, Giovanni
 ; APPLICANT: Rappuoli, Rino
 ; TITLE OF INVENTION: Helicobacter Pylori proteins Useful
 ; for Vaccines and Diagnostics
 ; NUMBER OF SEQUENCES: 7
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Chiron Corporation
 ; STREET: 4560 Horton Street
 ; CITY: Emeryville
 ; STATE: California
 ; COUNTRY: USA
 ; ZIP: 94608-2916
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/470,260
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/256,848
 ; FILING DATE: 21-OCT-1994
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: McClung, Barbara G.
 ; REGISTRATION NUMBER: 33,113
 ; REFERENCE/DOCKET NUMBER: 0316.001
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (510) 601-2708
 ; TELEFAX: (510) 655-3542
 ; INFORMATION FOR SEQ ID NO: 3:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1296 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-470-260-3

 Query Match 3.18; Score 206; DB 3; Length 1296;
 Best Local Similarity 20.18; Pred. No. 3.1e-06;
 Matches 261; Conservative 182; Mismatches 446; Indels 408; Gaps 70;

 QY 96 NQSDNK-DSOVIIVLVNGKEVKRQLVNDTRKGAAGFNR--NDVYKVNPAJENSMSG--F 150
 Db 72 NKTPDKPKVWKRI-----QAGKGFNEFPNKEDYLSLSLSSKIDGGWD 114

 QY 151 OGIIPLTYVKNENVLHREFSNDVKTGEGYV-----DFWSELMVPKDSFQKNGPLKQ 205
 Db 115 WGNARHWYVGGG---QNKLEYDKMDKAVGYTTLGSLNRFTGGDLV--NNQKATLRLGQ 169

 QY 206 FGLQPIINGQQY--YIDPTTGQPRKNFLQ--SGNNWYIFDSDTGVC-----TNALELQF 255
 Db 170 F-----NGNSTSYKSDADTRVDFNAKNKISIDFVFNINNVGSSGAGRASSTVLTLLQA 224

QY 256 AKGVTSSE-----OYRNGNRAYSDDKSIENVNG- 285
 Db 225 SEGITSKNAEISLYGATNLINLASSVVKLMGNVWMGRLOIVGAYLAPSYSTINTSKVTGE 284
 QY 286 ----YLPADTYWRPKQIKDKGTTWTDSKETDMRPILMVMWPNFLTQAYLYLNTM----- 334
 Db 285 VNFHLTVG----DKNAQAQI--IANKTINIGTL-----DLMQSAGLIIIAPEGGY 331

 QY 335 KQHCNLLPSALPFFNADADPAELNHYSEIVQONIEKRISETGNTDMLRULMDFVTNWP- 393
 Db 332 KQPNNTPTSGAKN-DKNESAKNDQESSQNNSTQVINPPNSAQKTEVQPTQVIDGPF 390
 QY 394 -----NWKNDSENVNPSGIFQGGFLKYENSOLTPYANSDYFLLGRMPINIKDKDTYRQ 447
 Db 391 AGGKDTVNNIRINTNADGTIRVGGF----KASLT--TNAAHHLHGKGGVNLNQ-A-SGR 443

 QY 448 EFLAN--DIDNSNPVVOAQLNLYLL-----NF-----GITANNO--QAN 487
 Db 444 SLIVENLTGNTVDGPLRVNQQVG--YALAGSSANFEKAGTETKNGTATFNNDISLGR 501
 QY 488 FDSVQVDAPONIDADLNIADQYFNAAYG-MOSDAYSNKHINILEDNHNADPEYFNKIGN 546
 Db 502 FVNLKYDAH---TANFKGI--DAGNGGFNIDFSGYDR-VNI-----N 539

 QY 547 POLTMDDTIKNSLNHGLSDATNRWGLDAIVHQSLADRENNSTENVVPIPNYSFVRAHDNNS 606
 Db 540 KLITASTNV-----AVKAFNINELIVK-----TNGISVGEYTHF-SEDIGS 579
 QY 607 QDQIONAIRDVTGKDYHTFTFEDEQKIDAYIODNSTVKKNLYNIPASY-----AI 659
 Db 580 QSRI-NIVRLETG--TRSLFSGGYKEKGEKLVIDE-----FYSPWNYFDRNKKV 629
 QY 660 LLTNKDTI-----PRVYGDLYDGGQYMBHQRYDYDTLNL-----KSRVK 702
 Db 630 EITNKLAFPGQSPWGTSKLNFNLTGQNAVMD-----YSQFSNLTIOGDEFNNQGTIN 684
 QY 703 YVAGQSMOTMSVG-----NNNLTFSVRYGKAM--TATDT----- 737
 Db 685 YLVGKGKATILSVGNAAAMFNNDIDSTGTYKPLINKSAQDLINKTEHVLKAKI 744
 QY 738 GTDETRTGGIGVYVSNTPNLKLVNDKVLHMGAAHKKQYRAAVLTITDGVINYISDQG 797
 Db 745 GNYSTGTNGISNV-----NLEEQKERLALY-----NNNRMDTCVVRNTDDI-----KA 789
 QY 798 APVAMTDENGDLYLSSHNLVYNGKEADTAVQGYANPEVSGYL--AWVPYVGSADNQDAR 855
 Db 790 CGMAIGDOS-----MVN-----NPDNTKYDIGKAWKNIGIS-----K 821
 QY 856 TAPSTEKN---SGNSAYRTNAAFSNVIFEAFSNFVYTPTKESERANVRIAQNADE- 908
 Db 822 TANGSKISVYVLGNSTPTENGNTNLTPTNTSN-----ARSANNAQAQNAQPS 873
 QY 909 -----FASLGFTSPMAPQYNSKD-----RPFLLSTIDNGY 940
 Db 874 ATPNLVAINQHDPGIIE-SVFELA---NRSKIDITLYANSGAQQGROLLQTLTIDSHDAGY 929
 QY 941 AFTDRYDLGMSFEPN---KYGTDEDLRNLQAAL-HK-AGLQVMADWVPDQIYNLPQKEVA 994
 Db 930 A---REKMDATASNEITKQLWATATTLNNTASLEHTSGQLTL----- 969
 QY 995 TVTRVDDRGNTWVKQAIINNLYVYNTIGGGEYQKKYGAGFLDKLQKLYPEIFTKKQVSTG 1054
 Db 970 -----SLSNAMILNSRL-----VNLRRHTNHTDSPAQRLOALKQKQKFALESAAE 1015
 QY 1035 VALDPSKQITSEWSAKYFNGTNILHRGSGYVLKADGGYVNLGTT-----TKQELPIQLTG 1109
 Db 1016 V-----LYQFAEKYEKPTNVW-----ANAIGGTSLNNGSNASLYGTSAGYDAYLNG 1061
 QY 1110 EKKQNGEGFVKGNGNYYFYDLAGNMVKNTEFIEDSVGNWYFFDQDGKVENKHFVDVDSY 1169
 Db 1062 QVEATVGGF--GSYGYSSFNKANSI-----NSGANNTNFVYSRIFANQHEFEFAQ 1112

```
QY 1170 GEKGTFFLKNGVSFRGLVOTUNGTYFYFDNYGKMYR 1206
Db 1113 GALSDD--OSSLNFXSALLQDLNQSXYHLAYSAAATR 1146

RESULT 19
US-08-471-491-3
; Sequence 3, Application US/08471491B
; Patent No. 6030611
; GENERAL INFORMATION:
; APPLICANT: Covacci, Antonello
; APPLICANT: Bugnoli, Massimo
; APPLICANT: Telford, John
; APPLICANT: Macchia, Giovanni
; APPLICANT: Rappuoli, Rino
; TITLE OF INVENTION: Helicobacter Pylori Proteins Useful For Vaccines And
; FILE REFERENCE: CHIR0044
; CURRENT APPLICATION NUMBER: US/08/471,491B
; CURRENT FILING DATE: 1995-06-06
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 1296
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-08-471-491-3

Query Match 3.1%; Score 206; DB 3; Length 1296;
Best local similarity 20.1%; Pred. No. 3.1e-06;
Matches 261; Conservative 182; Mismatches 446; Indels 408; Gaps 70;

QY 96 NQSDNK--DSOWIIVLVNGEVARQLVNDTKEGAAGFNR--NDVYKVNPAIENSSMSG--F 150
Db 72 NKTDPKDKVWRI-----QAGKGFNEFPNKEYDYLSLASKIDGGWD 114

QY 151 OGIIITIPVYKVENQVLVHRENDVKTGENTV-----DFWSELMPVKDSFOKNGGPKLQ 205
Db 115 WGNAAHYWVKGQ--QNKLEVDMDKAVGTITLSGLRFTGDDIDV--NMOKAHLRLQO 169

QY 206 FGLQVINGQY--YIDPTGQPRKNPLQ--SGNNWYFSDSTGVG-----TNALELQF 255
Db 170 F-----NGNSFTSYKDSADRTYVDFNAKNISIDNFEINNRRVNGSGAGRKASSTVLTQA 224

QY 256 AKGTYSNE-----QYRNGNAYSYDDKSIENWNG-- 285
Db 225 SEGITSDDKNAEISLDGATNLASSSVKLMGNWVRGLQYGYAYLAPSYSTINTSKVTGE 284

QY 286 ----YITADTWYRPKOILKDGTTWTDSDKTDMPRLVMWMPNTLQOAYLYNM----- 334
Db 285 VNFNHLTVG---DKNAAQAGI--IANKNTNIGTL-----DLAQSGALNIAPPEGY 331

QY 335 KQGNLLPSALFFPADAPPALNIHYSEIVQONIEKRISGTGNTDMLRMLMDFVFNRP- 393
Db 332 KDKPANTFSQGAKN--DKNESAKNPKQESSQNSNTQVINPPNSAQKTEVQPTQVIDGPF 390

QY 394 -----MWNKDSENVFSGIQFGQFLKYSNDLPYANSYRLLGRMPINKDQTYRQ 447
Db 391 AGGKTIVNINININADGTIRVGGF-----KASLT--TNAHLHTGKGVNLSNQ- 443

QY 448 EFLAN---DIDNSPVVQAEQLNLYLL-----NF-----GTIANND--QAN 487
Db 444 SLIVENLTGNTITVDGLRYNNQVGG--YALAGSSANFEKAGTDPKNGTATFNNDISLGR 501

QY 488 FSVRVDPDNTDADLMNTAQDYFNAAYG--WDSDAVSNKHINILEDMNHADPEYFNKIGN 546
Db 502 FVNLKYDAH---TANFKGI--DTGNGGFNTLDFSGVTDK-VNI-----N 539

QY 547 POLTMDDTIKSLNHLGSDATNRWGLDAIVHQSADRENSTENVYFNYSFVRAHDNNS 606
Db 540 KLITASTNV-----AVKNFNINELIVK-----TNGISVGTYTHF-SEDIGS 579
```

```
QY 607 QDQIONAIRDVTGKDYHTFTTDEQKIDAYTODQNSTVKKYNLYNPASY-----AI 659
Db 580 QSRI--NTRLETG--TRSLFSGGVKFKGKELVIDE-----FYSPWNYFDARNIKRV 629

QY 660 LITNKDTI-----PRVYVGDLYTDGGQYMEHQTRYDYDLTNLL-----KSRVK 702
Db 630 EITNKLAFGPGQSPMGTSKLMFNNTLQGNVMD-----YSOFSNLTQGGDFINNQTIN 684

QY 703 YVAGGQSMQTSVGG-----NNNLTLSVRYGKGAN---TATDT----- 737
Db 685 YLVKRGKVATLSVGNAAAMFNNDISATGYFKPLIKINSQADLIKKEHVLLKAKIIGY 744

QY 738 GDEPRTGIGVYVSWNTPNLKVNDKVVVLWNGAAHKHQYRAAVLTITDGVINYSQOG 797
Db 745 GNVSTGTNGISNV-----NLEQEFKERLALY-----NNNRMDTCVVRNTDDI---KA 789

QY 798 APVAMTDENGDIYLSHNLVNGKEADTAQGYANPDVSGYL--AVWVFGASDNQDAR 855
Db 790 CGMALGDQS-----MVN-----NPDNYKILIGKAMKNIGIS-----K 821

QY 856 TAPSTEKN---SGNSAYRTNAAFDSNVIFEAFSNFYVPTKESERARVRIQNADF--- 908
Db 822 TANGSKISVYILGNSTPTENGNTNLTPTNTSN-----ARSANALAQNAFFAQPS 873

QY 909 -----FASLGFTSFEMAPOYNSKD-----RTFLDSTIDNGY 940
Db 874 ATPNLVAIQNHDFGTIE-SVFELA---NRSKIDITLYANGAQQGRDLTLQTLIDSHDAGY 929

QY 941 AFTDRYDLGMSERN---KYGTDEDLRNAIOAL-HK-AGLQVMDADWPDQIYNLPGKEVA 994
Db 930 A---RKMDATARSANEIYQKLTATTTLNIAISLEHRTSGLOTL----- 969

QY 995 TVTRVDDRGNVKDAIINNLYVYVNTIGGGEYOKYKGGAFLDKLQKLYPELTIKKQVSTG 1054
Db 970 -----SLSNAMILNSRL-----VNLRRHTNHIDSFAKRLQALKDQKFASESAAE 1015

QY 1055 VALDPSOKITWENSAKYPNGTNIILHRGSGYVILKADGGQYVNI.GTT-----TKQFLPIQITG 1109
Db 1016 V-----LYOFAPKYEKPTNVW-----ANAGGTSLNNGSNASLYGTSAGVDAYING 1061

QY 1110 EKKQNGEYKVGNDGNYFYFDLAGNMVKNFTIEDSVGNWYFFDQDGKMNKHFVDVDSY 1169
Db 1062 QVEAIVGGF--GSYGYSSFNRRNSL-----NSGANNTFGYYSRIFANQHEFDFAQ 1112

QY 1170 GEKGTFFLKNGVSFRGLVOTDNGTYFYFDNYGKMYR 1206
Db 1113 GALSDD--OSSLNFXSALLQDLNQSXYHLAYSAAATR 1146
```

RESULT 20

```
US-08-466-662-3
; Sequence 3, Application US/08466662B
; Patent No. 6130059
; GENERAL INFORMATION:
; APPLICANT: Covacci, Antonello
; APPLICANT: Bugnoli, Massimo
; APPLICANT: Telford, John
; APPLICANT: Macchia, Giovanni
; APPLICANT: Rappuoli, Rino
; TITLE OF INVENTION: Helicobacter Pylori Proteins Useful For Vaccines And
; FILE REFERENCE: CHIR0057
; CURRENT APPLICATION NUMBER: US/08/466,662B
; CURRENT FILING DATE: 1995-06-06
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 1296
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-08-466-662-3
```


Db 1408 TLKAGNKKVKGKKNITFA----- 1427
QY 446 GQEFFLANDIDSNPVPQAEQLNWLYYLINFGTITANNQANFDSVRVDAPDNIDADLN 505
Db 1428 -----LANDSVKATVSDK-----LSLCT-----NCKN 1453
QY 506 IAQDPENAAAGDSDAVSNKHHINILEDHWHADPEYFNKIGNPOLWMDOTIKNS-----L 559
Db 1454 IYSDTKGLFAKDSKTGDANIHL-----NGIAS-----TLTDTLLNSGATTNLG 1498
QY 560 NHGLSDATNR-----WGLDAIVHQSLADRENNSTENNVIPNVSFVRADNNSQ 607
Db 1499 GNGITDNEKRAASKVDVLNAGNVRGKVPAS-----ANQVENI-----DFVATYD----- 1545
QY 608 DQIQNAIRDVTG-KOYHRTFFE--DEQKIDAYIQDNSTVKRYNLYNIPASYAILLNK 664
Db 1546 -----TVDFVSGDKDTSTVESKDKGKRTVKIGAKTSV IKDHN-----GKLFYTK 1592
QY 665 DTIPRVYGDLYTDGQYMEHCHTRYDYDTLTNLLKSRVKYA-----GQSQMOTMSVSGNNNI 721
Db 1593 E-----LKDANNNGVYVTE--TDQKDEBGLVLTAKAVIDAVNKAQVRVYKVTGANGQND 1644
QY 722 LRSVRYGKAMFATDTGT-----DEPTQIGVYVSNTPNLKLGVDNKKVVLHMGAAH 773
Db 1645 PATVASGTNVTADNGTGTAEYTKANDGSITVKYVNVKAD--GLKLD-GDKIVAD----- 1696
QY 774 KNOQTRAAVLTTDDGVINVTSDQAPVAMTDENGDLYLSSHNLVYNGKEADTA----- 827
Db 1697 -----TTLVLTADGKVT-----AP-----NNGD-----GKFKVDASGLADAL 1728
QY 828 -----VQGYANPDVSGYLAWVPVG-----ASDN-----OD 853
Db 1729 NKLSWTAAGKGTGSEVPANSGAGVEKAGDKVTKAGDNLK IKQSGKDFYSLKKEKD 1788
QY 854 ARTAPSTKNSGNSRYRNAAFDSNVIFEA-----FSNFVYPTPKESERANVRLAQ 904
Db 1789 LRSVERKANGGTGSESKITKDTGITPANGCAGAGANTANTI-SVTKDGISAGKAVT 1847
QY 905 N-----ADFFSLGTFVSEMAPOY-NSSKDRFTLD-----STDN----- 938
Db 1848 NVVSGLKFGDGHTLANGTVADFERHYDNVAKDLNLDKEKADNNTVADNTAAVVGDLR 1907
QY 939 GYAFTRDRLGSEPNKYGTDBLNAQALHAKGLQWADWVPQIYINLPCKEYATVTR 998
Db 1908 GLGWISADKTTGEPNQ-EYNAQVRANREVRKSGNGI-----NVSGKTLDNSTR 1956
QY 999 VDRGNVKKDAIINNGLYVWNTIGGGEYQKRYGGAFLDKLQKLYPEIFTKQVSTGVAID 1058
Db 1957 V-ITPELAKGEVKSNEFTVKNADGSETN-----LVKVGDMY-----ISKED-----ID 1999
QY 1059 P-SQKITEWSAKY-----FNGT-----NILHRGSGYVLKADGGY-----YNL 1095
Db 2000 PATSKPMGTGTEKYKVENGVKVSANGSKTEVTLTNKSGYV-----TGNQVADAIKSGFEL 2056
QY 1096 G-----TTTKQFLPT-----QLTGEKKOG-----NEGFVKGNDGNYIFY 1129
Db 2057 GLADRAEAEKAPAEAKDKQLSKDAETVNAHDKVPFANGLNLTKVSAAATVESTDAN----- 2112
QY 1130 DLAGNNKVNFTIEDSV-----GNNYFFDQDGKMYENKHFV--D 1165
Db 2113 -----GDKVTTTFVKTDVPLPLQIYNTDANGNKIVKADGKVELNADG-TASNKEVTLGN 2168
QY 1166 VDSYCKGTGYFLKNGVSGFRGLVQ-TDNGT--YFFDN-----YKMYRNQOTINAGA 1214
Db 2169 VDANGK-----VVKVTENGADKMYTTNADGAADKTKGEVSNKYSTDEK 2213
QY 1215 MYTILD-----ENGLIKASYNDAEY-PTSTD 1241
Db 2214 HVVRLDPNNSGKGVINDNANGVSEISATSTD 2245

RESULT 22
US-08-387-942C-3

; Sequence 3, Application US/08387942C
; Patent No. 5939289
; GENERAL INFORMATION:
; APPLICANT: BERTESVAG, HELGA
; APPLICANT: VALLJA, SVEIN
; APPLICANT: SKJAK-BRAEK, GUDMUND
; APPLICANT: LARSEN, BJORN
; TITLE OF INVENTION: DNA COMPOUNDS COMPRISING SEQUENCES
; TITLE OF INVENTION: ENCODING MANNURONAN C-5-EPIMERASE
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH, LLP
; STREET: P.O. BOX 747
; CITY: FALLS CHURCH
; STATE: VA
; COUNTRY: USA
; ZIP: 22042
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/387,942C
; FILING DATE: 09-MAY-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: MURPHY JR, GERALD M.
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 1809-106P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-205-8000
; TELEFAX: 703-205-8050
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1403 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-387-942C-3

Query Match 3.0%; Score 201; DB 2; Length 1403;
Best Local Similarity 20.1%; Pred. No. 8e-06;
Matches 304; Conservative 171; Mismatches 497; Indels 544; Gaps 79;

QY 25 DTTIKLNDQYQALNGLOVLLRPSKADGNPSGDNVTVDQFSKAYATTCGNFYV--KV 82
Db 72 ETVINKVDGWTQNTGMVR-----SAYGETSNFGMSDLTLDGNRDNLSAKV 118
QY 83 NGQVETSGWHATNQSNKDSQWIIVLVN-----GKEVKQLVNDT-KEGAAGFNND-- 134
Db 119 DG--WFNGYIPGQGDADRDVTLRVERREMSGYGDFEHTQINLTINDSVAHNDSLDGF 175
QY 135 --VTKVNPAB-----NSSMSGFQGIITLPTVTKVKNVQLVHRFSNDVKTCGEGNYDFWS 187
Db 176 VADYQVGVGVFENNYSYNDRHGFN--IVTSTNDFV-----LSNNVAYGNG----- 218
QY 188 ELMPVKDSFGKNGPLKQFGLQ-----TINCOQYYIDPTGQPRK-----NFLQSGNN 236
Db 219 -----GAGLVQRGSYDLPHPYD ILIDGAYVDNALEGVQLKMAHDVTLQ--NA 265
QY 237 WIYFSDSTGVGT-NALELOFAGKTGVSSNEQYRNGNAAY-----SYDDKSIENNVYLTA 289
Db 266 EIIYNGLYGVVRYGAQDVQILDNQIHNSQ-----NGAYAEVILQSYDDTAGVSGNFYVT 321
QY 290 DWTYRPK-----QILKDGTTWT-----DSKETDMRP ILMWVWPNITLQAYY 330
Db 322 GTWLEGNVISGSANSTYGIQPRADGTDYSSLYANS IDGVQGTG--AVRLYGANSTVSSQ- 377
QY 331 LNYMKQHGNNLPSLPFFNADAPALNHNHSEIVQOQIEKRISETGN----- 377
Db 378 -SGSQQAATLEGA-----GNDALSGTEAHTLTLGQAGDDRLNGDAGNDILDGAGGRNLT 432

QY 378 -----TDWLRT---LHDFVTPNPNWKNKSENPNFSGIQFQG---GFLKYEN 418
Db 433 GGAGADTFRFSARTSDSYRDSASFNDLITD---FDAEDSIDLSALGFTGLGDIY---N 485
QY 419 SDLTPYANDYRLLGRWPINIKDQYRGDFLLAND-----IDNSNPVQAELNWLXYL 473
Db 486 GTLLAKTNAEGT---RTYLSYBADAQRREFALDONFTGLFNDNLL-----531
QY 474 LNFGTTPANNQANFOSVRVADPNL-----DADLMN---IAQDFNAAAYGMS-----519
Db 532 --FDAAPATGTEGSDNLCGTADGETILVGGNDFTLNGGAGDDILVGGAGRDSLTCGAGAD 589
QY 520 ----DAVSNKHIILEDWHADEPEYKNGPOLTMDWDRIKNSLNH-GISDATNRRGLDA 574
Db 590 VRFEDLSDSQRYTTGGDQAD-----RILOFPTLDRIDVSALGFTGLGNGRN---GTLA 642
QY 575 IVHQSLADRENSTENVTPNKSFVRAHONNSDQIQNAIRDVTGKDYHFTFEDBQKGI 634
Db 643 VVLSAGDRDLDKSYDTDANGYSFELSLAGNYQGQL-----SACQFVATSQGGQ 592
QY 635 DAVIQ--DQNSTVKKYNLNIIPASYAIL-LTNKDTIPRYVYGDLYTDGGQYMEHOTRYTD 691
Db 693 MTLIEGTGNDT-----LAGTEANERLLGLDGRDNI--NGGAGDILLDGGAGR-----D 739
QY 692 TLTNLLKSRVYVAGGQSHQMTWSVGNNNILSVR---YCKGANTATDGTDETRETQ---745
Db 740 TLT-----GGT-----GADTFLESTRDSYRDSASFNDLITDFTDQDRI 780
QY 746 ---GIGVVVSNTPNLKLVNDKVLHMGAAHKNQYRAAVALTTTQDVINKYTSQQGAPVAM 802
Db 781 DLGLGLG--FSGFEN---GYDGTLLQVNA-----GTRYLKSFEA----816
QY 803 TDENGDLV-----LSHNLVV-----NGKERADTAVQ-----GYA 832
Db 817 -DANGOREFIALDGFSGQLDSCNVIPEPAVNAKDFGALGDGASDDRAIQAADAAYA 875
QY 833 NPDVSGYLAVPVPGASDNDQARTAPSTEKNSG-----NSAVRTNAAFDSNY-----880
Db 876 ---AGGVVILPAG-----EYRVSTGEGGCLMLKGVYLAGDGIQBTVKILDGSD 926
QY 881 -----FEARSNFVYTP-TKESERANVRIAQNADFFASIAFTSPMAPQYNS---925
Db 927 QKITGVMSAYGEETSNGHSDILTLGNNRDN-----YDFVMTNNVAYNGGAGLTIQRG 971
QY 926 SKDRTFLDSTID-----NGVAFTRDVLGHSBPNKYCTDEDLNATQALHAGLO-VH 977
Db 972 GQCGADRNTYIERVEIREMSGYGF-----DPHEQTINLTIHDSV---AHDNGLDGFPV 1020
QY 978 ADWVFDQIY-----NLPQREKAVATVRVDDRGVNWKDAILNNLYVVNTTGGGEYOKK 1029
Db 1021 ADYLVDSVFENNVAYNDRHGFNIYST-----YDFVMTNNVAYNGGAGLTIQRG 1071
QY 1030 Y-----GGAFIDKLQKLPETFTKKQVSTGVAIDPSQKITEKSAK-YFNGT---1074
Db 1072 SEDLAQPTDILIDGGAYYDNALE-----GVLFKMTNNVTLQNAEILYNGSSG 1118
QY 1075 -----NILHRS-----GYVLKA---1087
Db 1119 VRLYGTEDVQILDNIHNSQNGTYPEVLLQAFDSDSQVTGELYETNRIEGLNLDASDN 1178
QY 1086 -----DGQYYNL---GPTTKQFTLP IQLTG-----EKKQGNQ--FVKGND 1123
Db 1179 ANYAVRERDDGSDYTTLVNDISGGQVAVSLGSAHSLSGGTVEVPQGTGDNVLVGS 1238
QY 1124 GNYFY-----DLAGNWNKTFIEDSVGNWYFFDQDGKAVE 1159
Db 1239 ANDQYGGAGDRLDGGAGDLDLGGAGRDDLTTGCTGADTFVFAARTDSYRTDAG---VF 1295
QY 1160 NKHFVDVDSYKGTGYFFLKNQSVRSGGLVQDNDGTYTFYFDNKGKVRNQTINAGAMITL 1219
Db 1296 NDLIILDFDASEDR-----IDLALGFSG-----FGDGYNGTLLVQLSSAGTRTYL- 1340

QY 1220 DENGKLIKASYNDAE 1235
Db 1341 -----KSYEEDLE 1348

RESULT 23

US-09-514-302-2
; Sequence 2, Application US/09514302
; Patent No. 6338959
; GENERAL INFORMATION:
; APPLICANT: HATADA, Yuji
; APPLICANT: IGARASHI, Kazuaki
; APPLICANT: OZAKI, Katsuya
; APPLICANT: ARA, Katsutoshi
; APPLICANT: KAWAI, Shuji
; APPLICANT: ITO, Susumu
; TITLE OF INVENTION: GENE FOR ENZYME HAVING BOTH ALKALINE POLLULANASE AND
; TITLE OF INVENTION: ALKALINE ALPHA-AMYLASE ACTIVITIES
; FILE REFERENCE: 2173-105P
; CURRENT APPLICATION NUMBER: US/09/514,302
; CURRENT FILING DATE: 2000-02-28
; EARLIER APPLICATION NUMBER: 08/952,084
; EARLIER FILING DATE: 1997-11-10
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1938
; TYPE: PRT
; ORGANISM: Bacillus sp.
US-09-514-302-2

Query Match 3.0%; Score 200.5; DB 4; Length 1938;
Best Local Similarity 18.7%; Pred. No. 1.4e-05;

Matches 284; Conservative 201; Mismatches 508; Indels 523; Gaps 80;

QY 36 QALNQLQVLLRFSAACNCGDNTVTDQFSKNYATGG---NFDYVYKNGNOVFSGW 92
Db 7 RGMAGLLSILLIVSMFAGYLP-----ARAAETGDKRIEFSYERPDGN---YSGW 53

QY 93 HATNQSNDKDSQIILVYNGKEVKRQLVNDT--KGAAGFNENDYKYNPAIENSSMSG- 149
Db 54 NL-----W-----VWGTCVKDDQIDTFEKGKA-----YADIAVSDNADKVG 92

QY 150 -----FQGIITLPVYKNNVQLVHRFSNDVKTG-----EGNYVDVFW 187
Db 93 IIRKGDWEEKDFDGRSITIN-KIDNITKVHTSQEKFQIPDGSPPVVADGNADFEER 151

QY 188 ELMPYKDSFQKNGPLKQFGIQTNGQOYYIDPTTGQPRKNFLQSGNNWYFSDTGVG 247
Db 152 D---KELYAAGEMDKVEKVELSILGEYEMN---GEPEK-----ERFTYTLSDLPTG 197

QY 248 ----TNALELOFAGTSSNEQYRNGNAAYSYDDKSIENVNGYILTADTWYRPKOILK-DG 302
Db 198 EBHYTYLVTDQTEVTDPTVTDGRSVVEYVSDVQVSASFIPAKVDYQNVNAVVKDI 257

QY 303 TTWDSKEDMRPILMWPNLTQAYLYLNTYKQGNLLPSALPFFNADADPAELNHSE 362
Db 258 ESETETKIREMS-----INLSEIGGK-----EKATIDPA-LNELTV 292

QY 363 IVQONIE-----KRISGTNT-----DMLRTHLMHOFVTNTP 393
Db 293 AVKQGVYAGVKNLITAIDEFGRNHEGSATLEVQARTTGEKADFDWDQSVYV-FMLTDR 351

QY 394 MNWKSBNVNFSGIQGGFLKYENSDLTPYANSYR-----LLGRMPINIKDQYRG 446
Db 352 FFDGDSNNDPHGI-----GYDTSKGTQGGDFKGTORLBYLDDELGIN-----396

QY 447 QEFLLANDINSN-PVVOAQLNMLYLL-----NFGTITAN-NQANFDSVRVADPN 498
Db 397 -TIMISPVVDNIKEIDVRHSEGPDPYIAYHGYWADNFCELNPHFGSNADFHEM-IDAAHE 454

QY 499 IDADLM-NIAQDYFNAAYGM--DSDAVSN-----KHINILEDWNHAD-----P 538

Db 455 RGKINVDVVLNH--TGVLKPCDSSVANFFTDDEDRFCDKHLRGGSGEVRGELGLP 512
QY 539 EYFNKIGNPOL-----TMO-----DTI-----KNSLNHG 562
Db 513 DELTE--NPDVREQVYVQWQTDWIKSRKATKGTIDYFRVDYTKVHEDTTWAFKNAITKA 570
QY 563 LSD---ATNRWGLDVAHQSLADRENNSTENVIPNY-SFVRAHDNNSDQQTQNAIRDTV 618
Db 571 MPBKILIGAWG--ANVNDIAGYVLSGMDLSLDFKNYARDFANGQLDVAVQKLEARN 628
QY 619 GKDYHTFTFEDEOKGIDAYIQDNQSTVKYNYLNPASVAILITNNKDTIPRYVYGD--- 674
Db 629 SKLNNATIL--GQFLGSHDEDFEYVEVSDLGKQVAASLQITAKQ--PVIYVGEELGL 684
QY 675 -----LVT-----DGOYMEHQTRY--DPLTNLLSRVYKVVAGGSMQTM 713
Db 685 PKNDYPPYTNRNQMPWDDVGNLEIHYQKLLAFRNDNPNTFAKGRKKVAGSDS--- 740
QY 714 SVGGNNNLTISVYKGMATATDTGTDERTGIGVGVVSTPNLKLGVNDKVLVHNGAA- 772
Db 741 ----EGYLFSTYGENSY-----VGL-----NTEAAKD-VTLNFGSSE 776
QY 773 -----HKMQQYRAAVLTITDGVINYT--SDQGFVAMTDENGDLYLSSHNLVYNGKEE 823
Db 777 AVYTDYSGEYQA---NEEGOVTFIPAMEDGCVTLLEVENGAVPPVEBEPTGEIE 832
QY 824 ADTAVQGANPOVSGY-----LAWTPVGASQMDARTAPSTKNSGNSAYRTNAAPDSNV 879
Db 833 ENTLRIHQRTD--NSENIGLWLDV-----AAPSNNWPSGGTFFQAGNVTD--- 879
QY 880 IFEAFSNFYVTTKESERANVRIQADPFASLIFSPF-----MAPOYN----- 924
Db 880 ----YCAVYDVELAED-----AQNGHLV-LINTGDKGDKRAVELFSPDLNRINIK 927
QY 925 SKKRTFLDSTID-----NGYA----- 941
Db 928 QGSDEVFLYEPYDLPANTVRIHVERTNADYEGWGLWNWEDVESPDGFWNGAADAAGIGK 987
QY 942 FTDRVDLGMSE--PNKYG-----TDEDLNNAIQALHKAQLOVWADVWPOI 985
Db 988 YGAYDILKLEBANKIGFLFVNKQSGGQGTMDTMDLKQYNQLFYKGE-----EDKV 1039
QY 986 YNLPGKEVATVTRVDRGNWVKDAII-----NNNLXYVN 1019
Db 1040 YTNP---YGTVPALVSGEVLSDKLSLTFIRTEGLDLEELKEQLEIKNDVGDNDYSFTDV 1096
QY 1020 TIGG-----GEYQKK---YGAFLDKL-----OKLYPEIFT-KQVSTGVVAIDPSQK 1062
Db 1097 TIBGEKTVHVGFEFLKXIPFVSVTYLORTISVKSGLKIDENYAYDGLGABELHEDGTAT 1156
QY 1063 ITEWSAKYPNGTNILHRSGYVLKADGGQYVNLGFTTKOFLPIQLTCKKQGNEG----- 1117
Db 1157 LKWNSEKADNVSVLYD-----KVDQNEVD-----TIEMVKGDGQWNSVK 1197
QY 1118 FVKGN-----DGNIFYFDLAGNMYKNFPIE---DSVGNWYFFOOGKMKVKNHFVYDVS 1168
Db 1198 LTKONTGLDLSLAKGYHYHTEHGDVTNLALDPYAKSMAAWN--NEAGDKVGKAAIVDIGS 1255
QY 1169 YGEKGTFFLKNVSGFRGLVQTDNGTYFFDNYGKWNVQRTI--NAGAMIYTLDBN-OKL 1225
Db 1256 IGPE-----LOYADIPGFEKREDTIIYEVHVRDFTSDPNIGED 1293
QY 1226 IKASYNDAEYPTSYD 1241
Db 1294 LKAQGTFAFSEVKLD 1309

RESULT 24

US-08-968-685A-10

; Sequence 10, Application US/08968685A

; Patent No. 6214981

; GENERAL INFORMATION:

APPLICANT: TUCKER, KENNETH
APPLICANT: PLOSTILA, LAURA
TITLE OF INVENTION: MORAXELLA CATARRHALIS OUTER MEMBRANE
TITLE OF INVENTION: MORAXELLA CATARRHALIS OUTER MEMBRANE
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/968,685A
FILING DATE: No. 6214981ember 12, 1997
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Balgwin, Geraldine F.
REGISTRATION NUMBER: 31,232
REFERENCE/DOCKET NUMBER: 7969-060
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 2123 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-968-685A-10

Query Match 3.0%; Score 200.5; DB 4; Length 2123;
Best Local Similarity 18.9%; Pred. No. 1.6e-05;
Matches 266; Conservative 172; Mismatches 494; Indels 475; Gaps 58;

QY 49 SKADNGSGDNTYDQFSKNVATGTGNGFVYKNGVQVFGSHATNQSDKQSWIIV 108
Db 235 AKRASSIAGVNAQAIGFA---ATAVGSTQVNLNRGIALGFGSVQLQKND----- 283
QY 109 LVNGREYKRLVNDTKEGAAGFNRRNDVYKVNPAIENSSMSGFOGIIITLPVTKNENYLV 168
Db 284 -VNAANVRAYAPDD-----NQPIDNRYKATFKNGCATDVFSGNSN---- 322
QY 169 HRFND-----VKTGCONYVDFWSELMPVKDSFKQNGPLKQFGLQTINGQQYIIDPT- 221
Db 323 ---GNDSTIRRIIIVGAGSADTDVAVNAQLKAEAVRLANRQITFKGDDSNRVEKGLKTL 379
QY 222 --TGOPRKNELQSGNNWIYFDSGTGVTN--ALELQFAKTVS-----SNQYR 267
Db 380 TITGCAQTSAIT-----DHNIGVVGQGDGLKVQLAETLTSLKMYVTENI7ANEKVT 430
QY 268 NGNA-----AYSVDKSIENVNGYLTADTWYRP--KQILK-----DGTWTDSKE 310
Db 431 VGKRLATDKIGFTDNNGIDESKPYLDKDTGIHAGQKTKITAGVVDVDDAATYGQLKK 490
QY 311 TDMRPILWVWPNLTQAYLYNMYKQHCNLLPSALPFENADAPAEI-----NHYSEIVQ 365
Db 491 VNQT-----AESALQTFVTKVKDKNGN-----DANDSKIIITVCKNNKPDGTQ 532
QY 366 QNIEKRISSETG---NTDWLRLTLMHDFVTNPNMKNKDSNVNPFSCIOFGGFLKYENSEDLT 422
Db 533 VNTLKLKGVGVDTTETNGTGTFTGLNQNGL-----TVGNSTLNNDGLSVKNTNSN-- 584
QY 423 PYANSVRLGLRMPINIKDQTYRGQEFLLANDIIONSNEVYQAEOL-----NWLYYLLNFG 477

Db 585 -----KQIQVADSLITFTDISNKPAGIENTTILTRDGI GFANNY 626
QY 478 TITAN-----NDQANFDSVRVADPNIDADLMN-----TAQDY 510
Db 627 SLDAKPRLTPTGINAGGKBLTVQSAINPATNGGOLDPMRLSTANTKSGSAATIKDL 686
QY 511 FNAAY-----GHSDAVSNKHINILEDNHNADPEYFNKIGNPOLWMDTIKNSLHGLS 564
Db 687 YNLSQVPLFAGDTGNVYKLEILL-----KVGGKTTADDLTRKNIG-VVA 733
QY 565 DATNRGLDAIVHOSLADRENNSTENVIPNYSFVRAHDNNSODQIONAIRDVTGKDIET 624
Db 734 DSDDN-SLTVKLAKLTSLDLDAVNTKITL-----ASDKVTVDSCNNW----- 773
QY 625 FTEDEQKGDIAIQONSTVYKYNLYNPASVAIILLTKNDITPRVYGDLYTDGQGYME 684
Db 774 -----AKLQNGDLTFSKONTGATPA-----TNSKTI-----YG---VDGLKFTD 809
QY 685 HOTPRYDTLNLKSRVYV-----VLMGAALK-----NQYRAAVLTITDGVINYTSQOGAP 799
Db 918 KDFVSYDTVDFTNGNATPAKVYIDGKASKVAYVNDVGTTHLTGADGNKNOIGVKTIT 977
QY 800 VAMTDENGDLXLS-----SHNLVVGKEEADTA-----VQYANPDVSGYLAVWV 844
Db 978 LRTDAKDGKAINFSYNSGDDKALINAKDIADNLNTPLAGEIRNKTADTALQTPQVKV 1037
QY 845 PVGASNDQARAPSTPEKNSGNSAYETNAAFDSNVIFEAFSFFVYPTPKESERANVRIAQ 904
Db 1038 KENGDDNDADTI-----TVGKDAKTNOVNT-LKLKNGGLDIT 1076
QY 905 NADFPASLFTSEMAPQYNSKDRFTLSDTDNGVAFTRDVLGHEPNKYGTDBDLRN 964
Db 1077 NKDGTVFTGINT-----QSLKAGNNTLNN-----GLSIRKTAGNEQ----- 1115
QY 965 AICALHAGLOVADWVPQIYNLPGEVATVTRVDDRGVNWKDAIINNLLXVNTIGG 1024
Db 1116 -----IQVADGV-----KFAKVNQV----- 1133
QY 1025 EYQKKYGFAPLKLKLYPEITTKQV-----STGVAIDPSQKITEWSAKYFNGTNTILHRGS 1081
Db 1134 -----GAGIDGTTRI-----TRDEIGFAGTNGSIDSKPHLSKDGINAGGKITNIQS 1181
QY 1082 GYVLK-----ADGGQVYNEGTTKQPLPIQLGKKGQ-----NEGFKNGDGN-----Y 1127
Db 1182 GETAONSNDVAGGKIYDLKTELEN-----KLSSAKTAONSLSHEFSVADEQNNFTVSNPYS 1239
QY 1128 FVYDLAGNMVKNFTIEDS-----VGNWYFFDQDGKRVENKHFVDSYGEKGT 1175
Db 1240 SYDTSKTSVITFAGENGITTKVNGVVRVG-----IDQTKGLITPK-LTVGNNGKGLV 1293
QY 1176 FFLKNGVSRFGL-----VQDNGTYIFDNTGKAVRNOTINAGAMI-----YTLDEN 1222
Db 1294 IDSQONQITGLSNTLANVTNDKGSVTRTEOGKIITKEDKTRAASIVDWLSAGFNLOCN 1353
QY 1223 GKLIKASYNDAEYPTSTDVGKMLQDN 1249
Db 1354 GEAV-----DFVSTYTVNFADGN 1372

RESULT 25

US-09-268-347-36
; Sequence 36, Application US/09268347
; Patent No. 6335182
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.

; TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS
; FILE REFERENCE: 1038-860
; CURRENT APPLICATION NUMBER: US/09/268,347
; CURRENT FILING DATE: 1999-03-16
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 36
; LENGTH: 2411
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
; US-09-268-347-36

Query Match 3.0%; Score 200.5; DB 4; Length 2411;
Best Local Similarity 19.5%; Pred. No. 1.9e-05;
Matches 278; Conservative 152; Mismatches 439; Indels 557; Gaps 72;

QY 15 SLYNSAVSGFDFTIKLTNDYOALNGQLVLLRFSKAADGNPSGNDVYVDQFS--KNYAT 72
Db 1183 TLANGAAAGTD-----ASNGN-----TISVTKDGISAGNKEITNVKSAALKTYKD 1226
QY 73 TGGNFEDYVKNVGNQVEFSGWHATNOSNDKDSQWIIIVYNGKEVKRQLYNDTK--EGAAGF 130
Db 1227 TQNTAGATQPRANTAFA-----KODLVLTAPATGAAG-- 1260
QY 131 NRDVY-----YKVNPAIENSSMSGFQGIITLPVT 159
Db 1261 NGADAKAPDTTAATVGLRGLGWLSAKKTADETDQKEFHAAVKANAVEFVGKNGATVS 1320
QY 160 VKNENVQLVHRFSNDV--KTGEGNVYDFWSELMPVKUSFKQNGPLQKQPLQTTNGOY 216
Db 1321 AKTDN-NKSHVTIIDAFAKVGDDGLEK-----TDGKIKLVNDTGNLL 1365
QY 217 YIDPTGPKRNFLLQSGNNWYFDSGTGVTNALELQFAKGTVSSNEQYRNGNAYSVD 276
Db 1366 TVDANKAS----VAKGFENAVTTDTTAQGNANE----RGKVYVKGK--NGATATFTD 1415
QY 277 DKSINRVNGYLTADTWYRPKQILKDGTTWDSKETDMRPILMVMWPNLTITQAYILNYMKQ 336
Db 1416 KKKVATVGDVAKA-----INDAATFVKVENDSATI----- 1446
QY 337 HGNLPSALPFENADADPAELNHYSEIVQONTEKKISGTGNTDMLTLMHDFVT-----NN 392
Db 1447 -----DDSP-----TDDGANDALKA--GDTLTILKAGKN 1472
QY 393 PMWNKDSNVNFSGIFOGGELKYENSOLTPYANSDYRLGGRPINIKDQTYRGCEFFLA 452
Db 1473 LKVRKDGKNITFA----- 1487
QY 453 NDIDNSNPVQAEQLNWLYLLNPGTITANNQOANFSDVRVADPNIDADLMNTAODYFN 512
Db 1488 NDLVSKSATVSDK-----LSLGT--NGKNVNITS-----DTKGLNPAKD-- 1524
QY 513 AAYGMSDASVSNKHINILEDNHNADPEYFNKIGNPOLWMDTIKNS-----LNHGLSDA 566
Db 1525 SKTGDD-----ANHUNGAS-----TLTDTLLNSGATNLGGNGITDN 1563
QY 567 TNR-----WGLDAIVHOSLADRENNSTENVIPNYSFVRAHDNNSODQIONAI 614
Db 1564 EKKRAASVKDVLNAGMNVGVKPAKAS-----ANNQVENI-----DFVATYD-----TV 1605
QY 615 RDVTG-KDYHFTFE--DEQKGI DAYIQONSTVYKYNLYNPASVAIILLTKNDITPRVY 671
Db 1606 DFVSGDKDITSVTVESKDKGRTEVKIGAKTSVIKDH-----GKLEFGKE----- 1651
QY 672 YGDLVTDGGQWMEHOTRYDYDTLNLKSRVYV-----GGQSMQTSVSGNNILTSVRYG 728
Db 1652 LKADANNQVTTYE--TDGKDEGNGLVTPAKAVIDAVNKAQWVKTGTGANGQDDFATVASG 1709
QY 729 KGAMTADTGT-----DETRTQGIQGVVYSNTPNLKLVNDKQVWLHMGAAHKNQOYRA 780
Db 1710 TNVTFADGNGTTAEVTKANDGSIIVKYNVKAD--GLKLD-GDKIVAD-----T 1755

QY 781 AVLTTTGGVINTSDGAPVANTDENGDIYLSHNLVWNGKEADPAVOGYANPDVSGYL 840
Db 1756 TVLTVAGGVT-----AP-----NGD-----GKFPVDAS----- 1780
QY 841 AWVVPYGASD--NODARTAFSTEKNSG-----NSAYRTNAAFDSNVIFAFSNEVYTPTK 893
Db 1781 -----GLADALNKLSTWATAGKEGTGEVDPANSAQOEKAGD-KVTFKAGD----- 1825
QY 894 ESEANVRITQNA-DFFASL-----GFTSFENAPQYNSKORTFLDST--IDNGYAFTDR 945
Db 1826 -----NLKIKQSGKOPFTYSLKELKDLTSVE-----PKDANGGTGSSSTKITKDLGITPA 1876
QY 946 YDLGMEPNKYGTDEDLRNAIQALHKA-----GLQVMAOWPFDQIYNLPGEVA----- 994
Db 1877 NGAGACANTANTISVTKGISAGNKAVENTNVSGLAKFGDG-----HTLANGTVADFEKH 1931
QY 995 -----TVTRVDDRGNVWKDAIINNLLXYVNTIGG-----GEYQKKYGGAFLDK 1037
Db 1932 YDNAYKDLTMLDEKGADNPTVADNTAATVGLRGLGWISADKITGEPNQEYNQVYRNA 1991
QY 1038 LOKLYPEITFKKQVSTGVALDPQSKITWSAKYFNCTNIL--HRGSGYVLK-----A 1087
Db 1992 NE-----VKFKSGNGINV-----SGKTLNGTRVITTFELAKGEVYVKSNEFTVKNA 2035
QY 1088 DGGQ-----YYN-----LGTTKQFLP-----IQLTGEKKQ-----GNEG 1117
Db 2036 DGSETNLVKVGDWYISKEDIDPATSKPWTGKTEKYKVENGVVANGSKTEVTLTNKGSG 2095
QY 1118 FVKGNDGNYFYDLAGNMVYKNTF---IEDSVGNWYFFDQDGRKVE-----NKH--- 1162
Db 2096 YVTGN-----QVADAIKSGFELGLADAABAEKAFABESAKDKQLSKDKAETVNAHDKV 2148
QY 1163 -----FVDVDSYGKGYFFELKNGVSF-----RGLVQTDNGTY 1197
Db 2149 RFANGLNTKVSAAVTESTDANGDKVTTTFFVKTDVLPITQIYNTDANGNKIYKADGKWY 2208
QY 1198 FDRYKMRVROTNAGAMITYTLDENG-KLIKASYN-SDAEYPTSTD 1241
Db 2209 ELNADGTASNKVTLG-----NVDANGKKVVKVTENGAKWYTNAD 2250

Search completed: August 12, 2002, 07:53:55
Job time: 72 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 12, 2002, 07:52:43 ; Search time 29.2 seconds
(without alignments)
4116.704 Million cell updates/sec

Title: us-09-995-749a-2_COPY_531_1781
Perfect score: 6641
Sequence: 1 MSLTARPDVAAVPSLYNSA.....SDAEYPTDVGKMLDQMKL 1251

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database :

1: piri:.*
2: piri:.*
3: piri:.*
4: piri:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	2467	37.1	1508	2 T31098	probable dextran
2	2446	36.8	1431	2 A45866	dextranucrase (EC
3	2371	35.7	1449	2 T30857	glucosyltransferas
4	2371	35.7	1449	2 T30857	glucosyltransferas
5	2347.5	35.3	1375	2 JTO345	dextranucrase (EC
6	2276	34.3	1475	2 B33135	gtfB protein precu
7	2274	34.2	1592	2 A38175	glucosyltransferas
8	2250	33.9	1290	2 JC5473	dextranucrase (EC
9	2191.5	33.0	1365	2 A41483	glucosyltransferas
10	2175.5	32.8	1577	2 T30858	glucosyltransferas
11	2093	31.5	1518	2 A44811	glucosyltransferas
12	2005.5	30.2	1599	2 S22737	glucosyltransferas
13	317	4.8	2817	2 B97033	uncharacterized pr
14	264	4.0	2334	2 S32920	cell wall-associat
15	251.5	3.8	1385	2 D89824	hypothetical prote
16	250	3.8	2348	2 AD1841	hypothetical prote
17	246	3.7	2893	2 A64556	toxin-like outer m
18	241.5	3.6	601	2 S57962	cspC protein - Clo
19	241.5	3.6	1315	2 Y28679	fibrinogen-binding
20	239	3.6	2902	2 C71953	toxin-like outer m
21	229.5	3.5	3194	2 D71917	toxin-like outer m
22	225.5	3.4	2364	2 T40884	cytoxin L - Clos
23	222.5	3.4	1844	2 D71612	hypothetical prote
24	222	3.3	2366	2 S10317	toxin B - Clostrid
25	221.5	3.3	2020	2 C48399	ABC-type transport
26	221	3.3	1588	2 A86036	probable adhesin 2
27	221	3.3	1588	2 H91188	probable adhesin E
28	217.5	3.3	1806	2 AF1717	probable peptidogl
29	217.5	3.3	4936	2 A82515	hypothetical prote

30	216.5	3.3	1463	2 T30290	AAS surface protei
31	215	3.2	2340	2 B71704	cell surface antig
32	214	3.2	6713	2 B89921	hypothetical prote
33	213.5	3.2	1377	2 D90538	hypothetical prote
34	213.5	3.2	1530	2 A81396	peptidoglycan anch
35	213	3.2	1227	2 C97033	uncharacterized pr
36	213	3.2	2399	2 H71879	toxin-like outer m
37	212.5	3.2	1428	2 AC2224	hypothetical prote
38	211.5	3.2	1612	2 AD1517	probable peptidogl
39	211	3.2	1578	2 AD1517	peptidoglycan bon
40	211	3.2	2367	2 S70172	toxin B - Clostrid
41	210	3.2	1939	2 D97316	probable S-layer p
42	209	3.1	4182	2 T31102	filamentous hemag
43	209	3.1	4686	2 F82885	hypothetical prote
44	209	3.1	4919	2 T31105	hypothetical prote
45	207.5	3.1	583	2 S57721	cspB protein - Clo
46	207.5	3.1	1983	2 G86643	hypothetical prote
47	206.5	3.1	5005	2 F82884	hypothetical prote
48	206	3.1	1448	2 A12007	Subtilase family P
49	206	3.1	2044	2 A81180	probable peptidogl
50	206	3.1	2710	2 A37052	toxin A - Clostrid
51	205	3.1	2295	2 B71621	probable membrane
52	204.5	3.1	1475	2 A4765	alpha-amylase (EC
53	204.5	3.1	1582	2 AC1153	adhesin homolog lm
54	204.5	3.1	1999	2 B20118	hypothetical prote
55	204.5	3.1	2529	2 B64635	toxin-like outer m
56	203.5	3.1	1528	2 D85912	hypothetical prote
57	202.5	3.0	1819	2 D97033	uncharacterized pr
58	202.5	3.0	2167	2 AF1489	cell wall-associat
59	201.5	3.0	1217	2 F97177	alpha-glucosidase
60	201.5	3.0	1571	2 C91068	hypothetical prote
61	201	3.0	1403	2 S77624	mannuronan C-5-epi
62	201	3.0	1883	2 G82876	hypothetical prote
63	201	3.0	1902	2 B44858	lactocepin (EC 3.4
64	200	3.0	1645	2 F96907	phage-related prot
65	200	3.0	2500	2 G71609	hypothetical prote
66	200	3.0	2894	2 C64474	hypothetical prote
67	199.5	3.0	2013	2 AD1129	probable peptidogl
68	199	3.0	961	2 AD0548	putative autotransp
69	199	3.0	1902	2 B45764	lactocepin (EC 3.4
70	199	3.0	1902	2 S06997	lactocepin (EC 3.4
71	198.5	3.0	1643	2 D71630	outer membrane pro
72	198.5	3.0	1959	2 AG1085	hypothetical prote
73	198.5	3.0	3890	2 C89921	hypothetical prote
74	198	3.0	1132	2 H82887	hypothetical prote
75	198	3.0	1645	2 JN0896	crystalline surfac
76	198	3.0	4199	2 S76412	hypothetical prote
77	197	3.0	635	2 S57714	cspB protein - Clo
78	197	3.0	694	2 F97279	TPR-repeat-contain
79	197	3.0	1300	2 S07575	outer membrane pro
80	196	3.0	1144	2 A36968	PI-like adhesin pr
81	196	3.0	1203	2 S27545	pululanase - Ther
82	195.5	2.9	943	2 S31132	hypothetical prote
83	195	2.9	1238	2 A45396	hypothetical prote
84	195	2.9	1541	2 A37023	Iga-specific metal
85	194.5	2.9	1092	2 T30214	fibrinogen-binding
86	192	2.9	1881	2 H95076	zinc metalloprotei
87	192	2.9	2178	2 S55805	alpha-toxin - Clos
88	191.5	2.9	2154	2 F83068	hypothetical prote
89	191	2.9	918	2 G88545	protein F5982-12 f
90	191	2.9	1335	2 T30211	autolysin E - Stap
91	190.5	2.9	1268	2 B39789	hemagglutinin/memo
92	190.5	2.9	1270	2 E85649	hypothetical prote
93	190.5	2.9	1694	2 H64106	Iga-specific metal
94	190	2.9	1287	2 B53739	vacuolating cyto
95	190	2.9	1487	2 AG2560	hypothetical prote
96	189.5	2.9	1797	2 F69195	cell surface glyco
97	189.5	2.9	1946	2 A81449	hypothetical prote
98	189.5	2.9	2383	2 D64962	probable membrane
99	189.5	2.9	3283	2 AC1018	large repetitive p
100	189	2.8	1532	2 A36039	Iga-specific metal

ALIGNMENTS

```
RESULT 1
T31098
probable dextranucrase (EC 2.4.1.5), extracellular - Leuconostoc mesenteroides
C:Species: Leuconostoc mesenteroides
C>Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 11-May-2000
C:Accession: T31098
R:Monchois, V.; Renaud-Simeon, M.; Monsan, P.; Willemot, R.M.
FEMS Microbiol. Lett. 159, 307-315, 1998
A:Title: Cloning and sequencing of a gene coding for an extracellular dextranucrase (D8
A:Reference number: Z20981; MUID:98164374
A:Accession: T31098
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1508 <MON>
A:Cross-references: EMBL:AF030129; NID:g2766611; PID:g2766612; PIDN:AAB95453.1
A:Experimental source: strain NRRL B-1299
C:Genetics:
A:Gene: darB
C:Function:
A:Description: produces dextran composed only of alpha(1-6) glucosidic bonds
C:Keywords: glycosyltransferase; hexosyltransferase

Query Match 37.1%; Score 2467; DB 2; Length 1508;
Best Local Similarity 43.0%; Fred. No. 2.7e-116;
Matches 546; Conservative 174; Mismatches 423; Indels 128; Gaps 30;

QY 27 TIKLTNDYQALNGOVLRLFRSKAADNPSGDNTVTQFSKNYATTTGGNPFYVKNQ 86
DB 74 TIPVKNQD-----TQQIAANATQAEKVKASDTTDTQKQAEATNTNKEDSIDNLFQ 127
QY 87 VFESGHATNQSN--DKSQWITLV-----NKEVKRQLVNDTK 124
DB 128 LPAVPTANQTKYLEKDGKQKTYVTSNTLAKGLTVDNHNKQYFNNNGVQAKGFVTDNS 187
QY 125 EGAAGFNNDVYKVPAINSNKSGFGGIIPLPVTKENYVQLVHRFNDVTKGEGYVD 184
DB 188 K-----TYLDPNSGN-AVTGIQIGSTLAF-NDN-----GEQVAD 223
QY 185 FWSLMPVKDSF---QKNGPLKQFLOQTINGQQYYIDPTTQPRKNFLQSGNNWYFD 241
DB 224 FYT--APDGKTYFFDKGQATI---GLKAINGHNYFD-SLQLAKKGTGVIDGQVRYFD 277
QY 242 SDTGVGTNALEQFAKGVYSSNEQYRNGNAYSYDDKSTENWNGVLTADTWTYRQIILQ 301
DB 278 QESQGVSTTDSQIKEGLTSTQDTYAHNAVHSTDSADPDNPNGYLTASSWYRPRKDLRN 337
QY 302 GTTWTDSKETMRPILMVWPNLTLQAYLYNLMKQHGILLPSALPFENADADPAELNHY 361
DB 338 GQHWEATANDFRPIVSVWMPKQVQVNYLYMSQW--LIDNRQFSLKDMQAMLTAC 395
QY 362 ETVQONIEKRISETGNTDMLRPLMHDFVTNNPMWKNKSENVNFGIQGGFLKYENS DL 421
DB 396 TTVOQAETKIGVANSTAWLKTALIDDFIRTQPNWMSSEDP--KNDHLQNGALTFVNSPL 453
QY 422 TPYANGDYELLRMPIN-----IKQTYRQOEFLANDINDSNPVVQAEQLNWLTYLL 474
DB 454 TPDYNSFKLLNRTPTNGVPKYITDQSKGFELLLANDVDNSNPVQVQSEQLNWLHYLM 513
QY 475 NFGTITANDQANFVSVRVADPDNTDADLMNTAODYFNAAYGMD--SDAVSNKHINILEDW 533
DB 514 NFGSITANDSAANFGIRVDADVNDADLLQTAADYFRAAYGVKDNDAATQHLSILEDW 573
QY 534 NIADPEYFNKIGNPOLTMDDTTKNSLNHGLS-DATNRWGLDAIVHQSLADRENSTENVV 592
DB 574 SHNDPEYKDFGNLQTDVYHTQLINSLKDMRMKGTMQRFMDYLVNRRHDSNTENTA 633
QY 593 IPNYSFVRAHDNNSQDQONAIRDV--TGKDYHTTFFDEQKGDIDAYTQDQNSTVVKYNL 650
DB 634 IPNYSFVRAHDSVQTVIAQIISLHPVKNLSAPTAQDLAEAFKIYNNDKQADKKYQT 693
```

```
QY 651 YNPASAYALLTNKDTIPRVYVYGDLYTDGQYMEHQTRYDYLTLNLKSRVKYVAGGQSH 710
DB 694 YNMPASAYAMLLTNKDTIPRVYVYGDLYTDGQYMANKSYPFDALGLKSRKYVAGGQSH 753
QY 711 QTMSSVGGNNILTSVRYGKGMATADTGTDETRTGQIGGVVYNSNTNLKLVNDKVVLMHG 770
DB 754 AV----DQNDILTNRYGKGAMSVTDSGNADTRTQIGIGVIVSNKENLAKSGDVTVLMHG 809
QY 771 AAKKNOQYRAAVLTATTTDGVINYTSQGAPVAMTDEGDLVLSHNLVYNGKEADTAVQG 830
DB 810 AAKKNOQAFELLGTTADNLSTYDND-NAPVKYTNQDGLIFD-----NTEYIG 856
QY 831 YANPDVSGYLAVMVPVGSADNQDARTAPSTENKSNASAYRTNAAFDSNVIFEAFSNFYVT 890
DB 857 VRNPQVSGFLAVVVPVGSADSHQDARTLSDDTAHDGKTFHSAALDSQVYIEGFSNQAF 916
QY 891 PKESFRANVRIAQNADEFASLIGFTSFEMAFQYNSKDKRTFLDSTIDNGYAFTRDYLGM 950
DB 917 ATNTEDYTNNAVJAKNGQLFKDWGITFSQLAPQYRSSTDTSELDSTIQNGYAFTRDYLGY 976
QY 951 SPBNRYGTDEDLRNAIQALHKAGLOVMADWYDQIYNLPKGEVATVTRVDRGNVWKDAI 1010
DB 977 GTEKYGVTDQLRDAIKALHANGIAQIADWYDQIYNLPQGLATVTRTNSYGDRTNSD 1036
QY 1011 INNLYVNTTGGGYYQKYGAFDLKLYPELFTKQYSTGVVAIDPDSQKITIEMSAKY 1070
DB 1037 IDQSLYVYQSRGGYQAYGGAFLSDIQKKYPALFETKQISTGLPMDPSQKITIEMSAKY 1096
QY 1071 FNGTNILHRGSSVYLKADG-QQYNYL--GTTTKQFLPLQLTGEXKKQGNKGVKGNQNY 1127
DB 1097 FNGSMQKGGAGVILKDSGTQDYKYKVTNNNRDLPLKQLTDDLSE--TGFEVDNIGAVY 1154
QY 1128 FYDLAGNMVKNTFIEDSVGNVYFFDQCKGVENKHFVDVDSYGEKGTVEFLKNGVSFRGG 1187
DB 1155 -YTLGSLARNTFIQDNGNYYFYDSYGHLYTG--FQVINNH----HYFFLPNGIELVQS 1207
QY 1188 LVQ-TDNGTYTFDNGKVMRNQTI-----NAGAMIYT-----ENGK 1224
DB 1208 FLQNAAGSTIYFDQKGRQVFNQYITDGTATYATQNDGTMVTSGETIDGHKQFYKNGT 1267
QY 1225 LTKASYNSDAE 1235
DB 1268 QVKGQFVSDDT 1278

RESULT 2
A45866
dextranucrase (EC 2.4.1.5) precursor - Streptococcus mutans
C:Species: Streptococcus mutans
C>Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 07-May-1999
C:Accession: A45866
R:Honda, O.; Kato, C.; Kuramitsu, H.K.
J. Gen. Microbiol. 136, 2099-2105, 1990
A:Title: Nucleotide sequence of the Streptococcus mutans gtfD gene encoding the gluc
A:Reference number: A45866; MUID:91100958
A:Accession: A45866
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1431 <HON>
A:Cross-references: GB:M29296
C:Superfamily: cpl repeat homology
C:Keywords: glycosyltransferase; hexosyltransferase
F:181-201/Domain: cpl repeat homology <CP1>
F:1127-1146/Domain: cpl repeat homology <CP2>
F:1192-1211/Domain: cpl repeat homology <CP3>
F:1257-1276/Domain: cpl repeat homology <CP4>
F:1277-1297/Domain: cpl repeat homology <CP5>
F:1321-1340/Domain: cpl repeat homology <CP6>
F:1341-1361/Domain: cpl repeat homology <CP6>
F:1385-1404/Domain: cpl repeat homology <CP7>
```


Db 590 ISNSLDRSESGKSKEMANYAFVRAHSDSEVQSIQIILKNEINPOSTG---NTFTLDEM 646
QY 631 QKIDAYIQDNQSVTKKYNLYNPASAYAILLANKDPHPVYIGDLYTDGGQYMEHQTRY 690
Db 647 KKAPEYINKDMRSANKQYTYNIPSAIALMLTHKDTVPVYIGDMYTDGQYMAQKSPY 706
QY 691 DTLNLKSRVYVAGSQMOTMSV-GNNN-----ILTSRYGKGMATDTGTDTR 743
Db 707 DAETLLKGRIRYVAGGQDMKVNYIGYGTNCMDAAGVLTYSRYGTGANSASDTGTAE 766
QY 744 TQIGVYVNTNMLKGVNDKVVHLHGAHKKQYRAAALTTTDTGVINTSDQAP--VA 801
Db 767 NQGMAYVSNQPALRTSN--LTINMGAAHRRQAYRPLLTNDGVATYLNDSANGIVK 824
QY 802 MYDENGDLVSSNHLVVGKEADYAVQYANPDVSGVYLA VVPVVGASDNQDARTAPSTE 861
Db 825 YTDGNGNLTFSSANE-----IRGIRNPQVDGYLA VVPVVGASENDQYRVAPSKE 872
QY 862 KNSGSAIYRTNAAFDSNVIFEAFFSNFYVPTKESERANVRIAQNADEFASLGFTSPMAP 921
Db 873 KNSGLVYESNAALDSQVYIEGFSNFQDFVFPNPSQYTNKKIAENANLFKSWGITSFEAP 932
QY 922 QYNSKDRFTFLDSITDNGYATFDYDLGMSPEPNKYGTDEDLRNAIQALHKAQYVMA 981
Db 933 QYVSSDDGSLDSVIQNGYAFTRDYIDGMSKDNKYGLADLKAALKSLHAGVISAADW 992
QY 982 PQIYINLPKEVATVTRVDNRGNWMDKAIINNNLYVNT-IGGGEYQKYGGAFLDKLQ 1040
Db 993 PQIYINLPKEVATVTRVNNIGETKDGAIIDHSLIAAKRTFTFGNDYQKYGGAFLDK 1052
QY 1041 LYPEIFTKKQVSTGVAIDPQSKITFWSAKYFNCTNHLHRSYVYLKADGGQYVNLGTT-T 1099
Db 1053 LYPQIEDRVQISTCKRMTTDEKITKWSAKYMNSTNILDGRSEYVLKNGLNGY--GTNGG 1110

QY 1205 -----VRNQTINACAMITV-LDENGKLIKASINSDAE 1235
Db 1222 KQDPKPDNNNNSGRNQFVQIGNVWAAVYDGGKRVTHQNTNGQ 1266

RESULT 4
T30552
glucosyltransferase N - Streptococcus salivarius (fragment)
C:Species: Streptococcus salivarius
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
C:Accession: T30552
R.Jaffe, R.I.
submitted to the EMBL Data Library, February 1998
A:Description: Streptococcus salivarius V1477 gtfN.
A:Reference number: Z20854
A:Accession: T30552
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1449 <JAF>
A:Cross-references: EMBL:AF049609; NID:g2935545; PID:g2935546; PIDN:AAC05156.1
C:Genetics:
A:Gene: gtfN

Query Match 35.7%; Score 2371; DB 2; Length 1449;
Best Local Similarity 42.5%; Pred. No. 2.1e-113;
Matches 554; Conservative 170; Mismatches 147; Indels 164; Gaps 40;
QY 5 ARPDAVAVPSLYNSAVSGFD--TWIKLTNDQYQALNGQLQVLLRFSKAAQGNPFGDNTV 62
Db 52 AIPDITVDTGTVSNTTAQDPTTAAATND---VATDQATPAPFDLTDTIT-----NTV 104

QY 63 TDQFSKNYATTGNN-----FDYVKVNGQVFEFGHATNPQSNKDSQMIILVYNGKEVKR 117
Db 105 AANAVDTVATVGTDRAAATNDTATNDTAVD-----TNNNTTDTT---TTVYHRAATTE 156
QY 118 QLVNPKTKGAGFNDRNDYKYNPALIENSMGSGFOGIIILPVTYKNEANVQLVHRSNDVKT 177
Db 157 ERATGARKGPTGGR--ATPVNGTNNANNT-----VTVVNDLPA-----TNNVVT 201
QY 178 GEGNVDFWSELMPVKDSFQKNGPLKQFGLQTINGCOQYIDPTTGGQPRKNFLQSGNNW 237
Db 202 -----DGPSTH-----IKTINGKQYVVE-DGDTIRKNYVLERIGGS 235
QY 238 IYFSDTGTGVTNMLELQFAK--GTVSSNEQYR-----NG--NAAYSDDKSKTENVNGYLT 288
Db 236 QYFNAETGRLSNQKEVRFKDKNGTGSADSTNTVNVANGDKNAFYGTITDKDIELVDGTYT 295
QY 289 ADTWPRKQILKDGTTWTDSDKETDMRPILVWVMPNTLQAYLYNMYKQHGSLP SALPFF 348
Db 296 ANTWYPRKEILKDGKEMTASTENDKRPLLTVMWP SKAQASYLYNMYKEQGLGTQTYTSF 355
QY 349 NADA--DPAELNHYSEIYQONIEKRISSETGNTDMLRTLMHDFVTNNPMNKDSNVNPSG 406
Db 356 SSQTDQDAALAE-----VQKLEGRIAREGHTDMLRTTIKNFVKTPQGNSTSENLD--NN 409
QY 407 IQFGGFLKYENSOLFPIYANSDYRLLRMPIN-----IKDQTYRGQEFLLANDIDN 457
Db 410 DHLQGALLYNNDSRTSHANSDYRELLNRTPTSQTKHNP KYTKDTSNGGFEFLANDIDN 469
QY 458 SNPTVQAEQLNWLHYLLNFGCITANNDQANFDSVVDAPDNIDADMNIHQADYFNAAYGM 517
Db 470 SNPAVQAEQLNWLHYIMNIGITATGSEDENFQVGRVDAVDNVDNLLQIASDYFKA KYGA 529
QY 518 D--SDAVSNKHINILDMNHADEPFNKIGNPQLTMDDTIKNSLAHG--LSOATNFWGLDAI 575
Db 530 DQSQDQALKLSILEANSHNDAYNEDTKGQLPMDDPHMLALVYLLRPTIGNRSGVEPL 589
QY 576 VHQSLADRENNSHENVIPNYSFVRAHDNNSQDQIQNAIRD-----VTGKDHYHIFTFEDE 630
Db 590 ISNSLDRSESGKSKEMANYAFVRAHSDSEVQSIQIILKNEINPOSTG---NTFTLDEM 646
QY 631 QKIDAYIQDNQSVTKKYNLYNPASAYAILLANKDPHPVYIGDLYTDGGQYMEHQTRY 690
Db 647 KKAPEYINKDMRSANKQYTYNIPSAIALMLTHKDTVPVYIGDMYTDGQYMAQKSPY 706
QY 691 DTLNLKSRVYVAGSQMOTMSV-GNNN-----ILTSRYGKGMATDTGTDTR 743
Db 707 DAETLLKGRIRYVAGGQDMKVNYIGYGTNCMDAAGVLTYSRYGTGANSASDTGTAE 766
QY 744 TQIGVYVNTNMLKGVNDKVVHLHGAHKKQYRAAALTTTDTGVINTSDQAP--VA 801
Db 767 NQGMAYVSNQPALRTSN--LTINMGAAHRRQAYRPLLTNDGVATYLNDSANGIVK 824
QY 802 MYDENGDLVSSNHLVVGKEADYAVQYANPDVSGVYLA VVPVVGASDNQDARTAPSTE 861
Db 825 YTDGNGNLTFSSANE-----IRGIRNPQVDGYLA VVPVVGASENDQYRVAPSKE 872
QY 862 KNSGSAIYRTNAAFDSNVIFEAFFSNFYVPTKESERANVRIAQNADEFASLGFTSPMAP 921
Db 873 KNSGLVYESNAALDSQVYIEGFSNFQDFVFPNPSQYTNKKIAENANLFKSWGITSFEAP 932
QY 922 QYNSKDRFTFLDSITDNGYATFDYDLGMSPEPNKYGTDEDLRNAIQALHKAQYVMA 981
Db 933 QYVSSDDGSLDSVIQNGYAFTRDYIDGMSKDNKYGLADLKAALKSLHAGVISAADW 992
QY 982 PQIYINLPKEVATVTRVDNRGNWMDKAIINNNLYVNT-IGGGEYQKYGGAFLDKLQ 1040
Db 993 PQIYINLPKEVATVTRVNNIGETKDGAIIDHSLIAAKRTFTFGNDYQKYGGAFLDK 1052
QY 1041 LYPEIFTKKQVSTGVAIDPQSKITFWSAKYFNCTNHLHRSYVYLKADGGQYVNLGTT-T 1099
Db 1053 LYPQIEDRVQISTCKRMTTDEKITKWSAKYMNSTNILDGRSEYVLKNGLNGY--GTNGG 1110

QY 1100 KQFLPQLTGERKQKQEGFVKGNGN-----YFEDLAGNMVKNTEFSDSVGNWY 1149
 Db 1111 KVSLEP-KVVG-SNQSTNGDNGDGSKEKELFVSRYRYN-MGOYAKNAFTKDNQGVY 1167
 QY 1150 FPDQGGKVENKHFVDVDSYSGKGYFFLKNQVSRFGGLVQTDNG-TYEDNYGKW--- 1205
 Db 1168 YFDNSGRMAVGKTD---GKQ--YFFLANGVQLRDGYRQNRQGVFYDQNGVLSANG 1221
 QY 1206 -----RQGTINAGMIYT-IDENGKLIKASYNDAE 1235
 Db 1222 KQDPKPDNNNTSGRNQFVQIGNNVWAYDNGKRVIGHQNINGQ 1266

RESULT 5
 J0345
 deXtransucrase (EC 2.4.1.5) precursor - Streptococcus mutans (strain GS-5)
 N:Alternate names: sucrose 6-glucosyltransferase
 C:Species: Streptococcus mutans
 C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 24-Sep-1999
 C:Accession: J0345; C33135
 R:Veda, S.; Shiroza, T.; Kuramitsu, H. K.
 Gene 69, 101-109, 1988
 A:Title: Sequence analysis of the gtfC gene from Streptococcus mutans GS-5.
 A:Reference number: J0345; MUID:89137980
 A:Accession: J0345
 A:Molecule type: DNA
 A:Residues: 1-1375 <UED>
 A:Experimental source: GS-5
 R:Shiroza, T.; Veda, S.; Kuramitsu, H. K.
 J. Bacteriol. 169, 4263-4270, 1987
 A:Title: Sequence analysis of the gtfB gene from Streptococcus mutans.
 A:Reference number: A33135; MUID:87308013
 A:Accession: C33135
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-349 <SHI>
 A:Cross-references: GB:M17361
 C:Genetics:
 A:Gene: gtfC
 C:Function:
 A:Description: catalyzes the synthesis of both water-soluble and water-insoluble glucans
 C:Superfamily: cpl repeat homology
 C:Keywords: duplication; glycosyltransferase; hexosyltransferase
 F:1-34/Domain: signal sequence #status predicted <SIG>
 F:35-1375/Product: glucosyltransferase #status predicted <MAT>
 F:1126-1145/Domain: cpl repeat homology <CP1>
 F:1253-1272/Domain: cpl repeat homology <CP2>
 F:1318-1337/Domain: cpl repeat homology <CP3>

Query Match 35.38; Score 2347.5; DB 2; Length 1375;
 Best Local Similarity 41.38; Pred. No. 3.1e-112;
 Matches 538; Conservative 181; Mismatches 408; Indels 175; Gaps 35;

QY 19 SAVSGFDITKLTNDQYQALGQLVLRFSKAADG--NPSGDNTVTDQFSKNVATTGNN 76
 Db 29 TSLGSLVKADSTDRQAVYFSQASLVTTSEAKETLTATDTSTATSQPTATVDN 88
 QY 77 FDIYVKNQVDFSGHWTNQSNDKDSOMIIVLVNGKVKRQLVNDTKEGAGFNRNDVY 136
 Db 89 VS-----TTNQS-----TNTNTA-----NPNV 107
 QY 137 KYNPALENSMSGFQGIITLPVTKENVQLVHRSNDVKTGEGHYV-----DFW 186
 Db 108 KETTSEQAKTNSDKIITTSKAVNR-----LTATGKFPVANNNTAHPKTVT 154
 QY 187 SELMPVQDSQKNGPLKQ-----FGLQTINGQOYIIDPTGQPKNFLL 231
 Db 155 DRIVPKPI-----GKLQPSLSQDDIAALGNVKNIKVNGKYYIY-KEDGTQKNVAL 209
 QY 232 QSGNNIVFDSGTGVTNALEIQFAGTVSSNE---OYRNGNAAYSDDKSIENVNGYLT 288
 Db 210 NINGKTFEPFDGALSNTLPSEK--KGNITNDNTNSFAQYNOVYSTDVANFEHVDHILT 267

RESULT 6

B33135

gtfB protein precursor - Streptococcus mutans

QY 289 ADTWYRPKQILKDGTTWDSKETDMRPILAMVWPNLTQAYVILAYMKOHGNLLPSALPFF 348
 Db 268 AESWYRPKYILKDKTWTQSTKDFKPELLMTWMPDQETQROYVVMNAQLGIHQT----Y 323
 QY 349 MADAPAELENIHSEIVQONIEKRISETGNTDWTLYMDFVTNMPMKNKSDENVNFSGIQ 408
 Db 324 NTATSPGLNLAAQIOTKIEKTAENTNLRQTIISAFVKTQSAMNSDSEKFPDHLQ 383
 QY 409 FQGGFLKYENSGLTPYANSDYRLJGRMPIN-----IKDOTYRGOERLLANDINSN 459
 Db 384 -KGALLYSNNSKLTQSQANSYILNRTFTNOTGKDPRTADRTIGGTFEFLANDVDSN 442
 QY 460 PVVQAEQLMWLYLLNFCETITANNDQANFDSVRVADPDNIDADLNLIAQDYFNAYGN-D 518
 Db 443 PVVQAEQLMWLHFLMNFQNIYANDPDANFDSIRVDADVNDVADLLQIAGDYLKAAGTKH 502
 QY 519 SDAYSNKIINILEDNHADPEYFNKIGNPQLTMDTITKNSLNHGLSDATN-RWGLDAIVH 577
 Db 503 NDKAANDHLSLEAWSYNDTPYLHDDGDMNMINNRNLSLLYSLAKPLNQSGMNPILT 562
 QY 578 QSLADRENNSTENVVIPNYSFVRAHDNNSQDQIONAIR-----DVTGKDYHTFTFEDBOK 632
 Db 563 MSLVNETDQNAETAAPVPSYSIRAHDSVQDLIRNIIRTEINPNVVG---YSFTTEELK 619
 QY 633 GIDAYIQDNSTVKKYNYLNIIPASVAILLTNKTIPRVYIGDLYTDGQYMEHQTXYDT 692
 Db 620 AFETYNKDLLATEKKYTHYNTALSVALLLTNKSSVPRVYVYGDMEFDDQYMAHKTINYE 679
 QY 693 LTNLLKSVKVVAGGQSMQTSVSGNNILTSVRVKGAMTATGTDDETTQIGVYVS 752
 Db 680 IETLLKARIKTVSGQAMRNOQV-GNSELIISVRVKGALKRATDGTDRTTSTGVAVLEG 738
 QY 753 NTPNLIKGVNDKVVLMHGAHKNQYRAAVLTITDGVINYTSDDQAP--VAMTDENGPLY 810
 Db 739 NNPSELRKASDRVVVNGAAHKNQAYRPLLTDTNGIKAYHSDQEAAGLVRYTNDREGLI 798
 QY 811 LSSHNLVNGKEADTAVQGYANPDVSGYLAVVPGVGSADNODARTAPSTKKNSGNSAYR 870
 Db 799 FTA-----AD--IKGYANPQVSGYLVGVVPGVGAADQDVRVAASTAPSTDGKSVH 846
 QY 871 TNAAFDSNVIFEASNFVYTTKESERANVRIAQNAFASLGIFSEMAPQYSSKDRPT 930
 Db 847 QRAALDSRVNMFEGFSNFQAFATKKEEYTNVVIKNNVDFAEWGYTDFEMAPQYSSYSDGS 906
 QY 931 FLDSIDNGYAFIDRYDLGMSPEPKYGTDEDLRNAIQALHAKGLQVMADWVDDQIYNLPG 990
 Db 907 FLDSVIQNGYAFIDRYDLGISPKNKYGTADDLVKAIKALHKSGLKVMADWVDDQYALPE 966
 QY 991 KEVATVTRVDGNGVWKKDAIINNLIVYV--TIGGGEYQKKYGGAFDLKQKLYPEIFTKK 1049
 Db 967 KEVVTATRVDXKYGTPVAGSQIKNTLYVVDGKSSGKQDQAKYGGAFLELQAKYPFLPARK 1026
 QY 1050 QVSTGVADIPDSQKITENSAKYFNGTNIHLRGSYVLKADGGYYNLGTTTKQFLPIQLTG 1109
 Db 1027 QISTGVPMDPSPVKIKQWSAKYFNGTNIHLRGSYVLKADGGYYNLGTTTKQFLPIQLTG 1085
 QY 1110 EKKQGNFVKG--NDG--NYIFYDLAGNMVKNTEFSDSVG-NWYFFDDQKRWENKHFVD 1165
 Db 1086 NPNHGTSSSTGLVFDGKGYVYSTSGNQAKNAFI--SLGNWYTFEDNNGYMTVGAQSIN 1143
 QY 1166 VDSYGERGTYEFLKNGVSRGGLVQT-----DNGTY-----YFDNYGK 1203
 Db 1144 -----GANYYFLSNGIQLRNAIYDNGKNVLSYTGNDGRYENGYYLFGQWRVYFQN-GI 1196
 QY 1204 MVRMOTINAGMIYTLDENGKLIKASYNDAEYPTSTDVGNK 1245
 Db 1197 MAVGLTRVHGAVQY-FDASG-----FOAKGQFITAD-GKL 1230

C:Species: Streptococcus mutans
C:Date: 23-Oct-1990 #sequence_revision 23-Oct-1990 #text_change 15-Oct-1999
C:Accession: B33135; A33128
R:Shiroza, T.; Ueda, S.; Kuramitsu, H.K.
J. Bacteriol. 169, 4263-4270, 1987
A:Title: Sequence analysis of the *gtfB* gene from Streptococcus mutans.
A:Reference number: A33135; MUID:87308013
A:Accession: B33135
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1475 <SH1>
A:Cross-references: GB:M17361; NID:g153639; PIDN:AAA88588.1; PID:g153640
R:Shiroza, T.; Ueda, S.; Kuramitsu, H.K.
Submitted to the Protein Sequence Database, September 1990
A:Reference number: A33128
A:Accession: A33128
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-171, 173-641, 'N', 643-1475 <SH2>
A:Experimental source: strain GS-5
C:Superfamily: cpl repeat homology
F:1096-1115/Domain: cpl repeat homology <CP1>
F:1224-1243/Domain: cpl repeat homology <CP2>
F:1289-1308/Domain: cpl repeat homology <CP3>
F:1354-1373/Domain: cpl repeat homology <CP4>
F:1419-1438/Domain: cpl repeat homology <CP5>

Query Match 34.3%; Score 2276; DB 2; Length 1475;
Best Local Similarity 42.4%; Pred. No. 1.6e-108;
Matches 518; Conservative 183; Mismatches 382; Indels 140; Gaps 37;

QY 107 IVLVGKVKRQLVND-----TKGAGFENDVYKYPALENSMS 147
DB 35 LVKADSNESKQISDNTSVVTANEESNVITEATSKQEAASQTN--HTVTTSSSTSV 92
QY 148 SGFGIITLPVVKVENVQLVHRFSDNVKIGBNVDFWSEL-----MPYKD 194
DB 93 VNPKEVSNPIV-GETASNGEKLQNTT-----VDKTEAANNISKOTTEADTDVD 146
QY 195 SFQKGNGLP--KQFGLQTNGQYVYIDPTQPRNFLOSNNIYFD-----SDTVGV 247
DB 147 DSNAAANLQLEKLPVVKETDGGKYTYD--NNGKVRNTFTLIADGKILHFDGTGAYTSD 205
QY 248 TNALEQFAKGTVSSNEQYRNGNAAVSDYDKSIENGVLTADTATVRKQILKCGTWTWD 307
DB 206 TVNKDLY-----TTRSN-LYKKNQVTDRAQSFHVDHLYTAESWYRKYILKDKGTWTQ 260
QY 308 SKETDMRPLVWVWPTLTQAYLYNMQKRGHLLPSALPFFNADAPALNHYSIVQON 367
DB 261 STEKDFRPLMTWPDQETQRYVYVNMNAQLGINKT---YDDTSNQLNLNIAAATQAK 316
QY 368 IEKRISETGNTDHLTLMHDFVNNPWNKSDENVFSGIQFGGFLKYEN-SDLTPTAN 426
DB 317 IEAKITLTKDNLQRTISAFYKTSOANSDSEK-PFDD-HLQNGAVYDNEGKLTPTAN 374
QY 427 SDVRLGRPIN-----IKDQYRQDFLLANDIDNSNVPVQAOLNLYLLNG 477
DB 375 SNVRLNRPPTQTKKDPYADNTIGYEFLLANDVNSNVPVQAOLNHLFLMNG 434
QY 478 TITANDQANFDSVRVDPDNDIDADIMTAQDYFNAAVGM-DSDAVSKNHILEWNEHA 536
DB 435 NIYANDPDANFDSIRVDVNDVADLQIAGDYLRKAQIKHNDKRAANDHLSILERSDN 494
QY 537 DPEYFNKIGNPQMTODTATKNSLNHGLSDATN-RWGLDAIVHQSLADRENNTENVVIPN 595
DB 495 DTPYLDGDDGNMINDMKRLSLFLSLAKPLQSRSGMNPILTNISLVNRTDDNAETAAPS 554
QY 596 YSFVRAHDNNSQDQIONAIR-----DVTGKDYHTTFEQRKIDAYIQDQNSVKKYNL 650
DB 555 YSFIRAHDSKVODLIADIITKAENPNVVG---YSFTMEIRKAFENKDLATEKRYTH 611
QY 651 YNPASAYAILTNKOTIPRVYTGDLTYDGGQMEHQTRVYDTLTNLLKSRVYVAGQSM 710

DB 612 YNTALSALLTNKSSPRVYVYDGFDDGQYMAHKRTINYEATITLTKARIKYVSGQAM 671
QY 711 QTMVGGNNILTSVRYGKGGAMTATDTGTRTQGTGVVVSNPNLKLGVDNRKVLHMG 770
DB 672 RNOQV-ENSEITSVRYGKALKATDGTDRTRPSGAVIDGNPNPSLKLKASDRVYVNM 730
QY 771 AAHNQYRAAVLTDTGVTNYTSDQAP--VAMTDENGDLYLSSHNLVYNGKEADTAV 828
DB 731 AAHNQAYRPLLTDTNGIKAYHSDQEAAGLVRYTNRGELIFTA-----AD--1 778
QY 829 QGYANPDVSGYLAIVVVPVGAS--DNQDARAPSTKNSGNSAYRNAFDSNVIFEAFSN 886
DB 779 KGYANPDVSGYLAIVVVPVGAAALIKMFALRLA---RPHQOMASVHQNAALDSRVNMFESN 835
QY 887 FVYPTPTKESRANVRIAGNADFFASLGFSTFEMAPQYNSSKDRFTLSDTNGYAFTRDY 946
DB 836 FQAFATKKEEYTNVIAKNVDKFAEWGVTDFEMAPQVSSDTGSLDSVQNGYAFTRDY 895
QY 947 DLGMSPNKYCTDELRNAIQALHKAQLQVADWVVPQIYNLPKEVATVTVRDDRNVW 1006
DB 896 DLGISKPNKYCTADDLVKAIRALSHKGIKVNADWVVPQMYAFPEKEVTVATRVKYGTPV 955
QY 1007 KDAIINNLLVYN-TIGGGEVQKYGGAFLDKLQKLYPEIFETKKOVSTGVAIDPSQKITE 1065
DB 956 AGSQIKNTLYVVDGKSSGKDDQAKYGGAFLEELQAKYFELFARKQISTGVPMDFSVKIQ 1015
QY 1066 WSAKYFNGTINILHRGSGYVLKADG-GYYNLTGTTK-QFLPIQLTGEKKQEGEVKGN 1123
DB 1016 WSAKYFNGTINILHRGAGYVLKQDAINTVFNISDNKEINFLPKTLNQDSQ--VGFSDYDK 1073
QY 1124 GNYFYDLAGNNVKNFTSDSVGNWYFDDQDKVKNKHFFVDVDSYGEKGYFFLKNQVS 1183
DB 1074 G-YVYTSSTGQAKNTFISEG-DKWIYFDNNGYMTGAOSIN-----GVNYEFLSGLQ 1125
QY 1184 FRGLVQPDNGTY-YFDNYGKVRN-----QTINAGAM-----TYTLDEN 1222
DB 1126 LRDAIKNEDCTIAYGNDGRRYENGYYQFMSGVWRHFNNGEMSVGLTVIDGQYQYFDEM 1185
QY 1223 GLKLIKASYNDAEYPTSTDVGMK 1245
DB 1186 G-----YQAKGKFEVTTAD-GKI 1201

RESULT 7
A38175
glucosyltransferase precursor - Streptococcus sobrinus
C:Species: Streptococcus sobrinus
C:Date: 28-Aug-1992 #sequence_revision 28-Aug-1992 #text_change 15-Oct-1999
C:Accession: A38175
R:Abou, H.; Mtsamura, T.; Kodama, T.; Ohta, H.; Fukui, K.; Kato, K.; Kagawa, H.
J. Bacteriol. 173, 989-996, 1991
A:Title: Peptide sequences for sucrose splitting and glucan binding within Streptococcus
A:Reference number: A38175; MUID:91123227
A:Accession: A38175
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1592 <A80>
A:Cross-references: GB:D90213; NID:g217032; PIDN:BAAL4241.1; PID:d1014946; PID:g2170
C:Superfamily: cpl repeat homology
F:1093-1112/Domain: cpl repeat homology <CP1>
F:1222-1241/Domain: cpl repeat homology <CP2>
F:1287-1306/Domain: cpl repeat homology <CP3>
F:1330-1351/Domain: cpl repeat homology <CP4>
F:1352-1371/Domain: cpl repeat homology <CP5>
F:1402-1420/Domain: cpl repeat homology <CP6>
F:1465-1484/Domain: cpl repeat homology <CP7>
F:1513-1532/Domain: cpl repeat homology <CP8>

Query Match 34.2%; Score 2274; DB 2; Length 1592;
Best Local Similarity 45.9%; Pred. No. 2.2e-108;
Matches 496; Conservative 167; Mismatches 340; Indels 78; Gaps 29;

```
QY 158 VYVKNENVLVHRSNDVKTGBGNVYDFHSELMPYKDSFKQNGNGLKQFGLQTINGQQYY 217
Db 111 :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 116 VPTENEN-----QGFDTDELAELAKNATAESDIP-----SDLAKMSNVKQVDGKYY 163
QY 218 IDPTTGOPKKNLLOSGNNWYFDSYGV--GTNALELOFAGKGTVSSNEQ-YRNGNAAYS 274
Db 164 YD-QDGNVKNFAVSGDKIYFD-ETGAYKOTSVDADKSSVASVONATIFAANNRAYS 221
QY 275 YDDKSIENGYLTADTYRPOKILKDGTTWDSKETDMPILMVMWPTLTQAYLYNTM 334
Db 222 :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 222 TSAKNFEAVDNYLTADSWYRPSILKDGKTWESGKDDRPLMLMAMPDTETKRYNYVM 281
QY 335 KQHGNLPSALPPFNADAPAEHLNHYSEIVQONIEKRISGTNTHLRLMHDFTVNNPM 394
Db 282 :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 282 ----NRVVGIDKTYTAETSCADITAAAELOVQAEIKGITSENNKLWREAIASAFVKTQ 337
QY 395 WNKDSNVNFSGTOFGGFLKYEN-SDIAPYANSYDVRLLGRMPINIKDQ-----TYR--- 445
Db 338 WGESEKPYDD--HLONGALLFDNQDLDLPDQOSNVRLNRTPTNGTSGLDSEFFYNPD 395
QY 446 ---GOEFLANDIDNSPVVQAEQLNWLNYLLNFGTITAMNDQANFDSVRVDAPDNIDAD 502
Db 396 PLGGYDELLANDVDNSPVVQAEQLNWLNYLLNFGSIYANDADANFDSIRVDAEDNVDAD 455
QY 503 LKNIAODYNNAAGMD-SDAVSNKHINILEDNNHDPPEYFNKIGNPQLTMDDTIKNSLKH 561
Db 456 QGIGSSDYLKAAVIGIDKNKNNAHNSIVSEANSNDTPYLHDDGDMLMNDKFRLSMLW 515
QY 562 GLSDATN-RWGLDAIVHQSADRENSTENVIIPNYSPVRAHNNNSDQIQONAIR-DVWG 619
Db 516 SLAKPTDVRSGLNPLHNSLVREVDREVEIVPSVSPARAHNDSEVQDIIRDIKAEINP 575
QY 620 KQY-HPTTEDEKGDIDAYIQDQNSTVTKYKYNLYNPASVAILLTKNDKTPRIVYGLD 678
Db 576 NSPGYSFTEIDQAEKIXNEDLKKDKKTYTHNVPLSYTLTLTKNGSGIPRVYVGMDFD 635
QY 679 GGOYMEHQTRYDTLNLKSRVYKYGQSGQMTMSVGGNNNTLTSVRYGKGNWATDGTG 738
Db 636 DGOYMAKNTVYDAIESLLKARKMYVAGQAMQNYQI-NGGEILTSVRYGKALKOSDKG 694
QY 739 TDBTRQGGIVGVVSNTPNLKLGVDKRW-LHMGAAHKNQYRAAVLTITDGVNITYSDQG 797
Db 695 DATTRSGGVVGMVGNQPNESL--DGRVVALNMGAAHANQYRALVYSTKDGVAITYATDAD 752
QY 798 AP-----VAMDENDGLYLASHNLVYNGKEEATATAGVGYANPDVSGVILAVVPGASDNQD 853
Db 753 ASKAGLVKKTDENGLYLFNLDDL-----KGVANPQVSGFLQYVWPVGAADDQD 800
QY 854 ARTAPSTKNSGNSAVRTNAADPSNIVIEAFNSFVTPTKESPRANVIAQNAADFEASLG 913
Db 801 IRVAASDTASTDGLSHQDAMDSRVHFEFSNFQSFATKEEYTNVIVANNVDFKFSWG 860
QY 914 FTSFEMAPQNSKRTFLDSTYDNGYAFATDRTYDLGMSFPNKYGTDEDLRNAIQALHAKG 973
Db 861 ITDFEMAPQVSTDGOFDSDVIONGYAFATDRTYDLGMSRANKYGTADQLVKAIALKNAKG 920
QY 974 LQYMAWDPQIYINLPCKEVAITYRVDGRGNWKKDAIINNLYVWAT-TGGGYKQKYG 1032
Db 921 LKYNMADWPDMQYTFPKQEVVYTRTDKFKPTAGSQINHSLYVVTDTKSSGGDYQAKYG 980
QY 1033 AFDLKLQKLYPEFTFKKQVSTGVAIDPSOKITSEWSAKYFNGTNILHRSGSYVLKAD-GGQ 1091
Db 981 AFLDELKEKPELFTKQMTGTGAIDPSVKIKOWSAKYFNGSNILRGADYVLSQDQVSNK 1040
QY 1092 YYNLGTTTKQFLQTLTGERKQNGEFGVKNQDNYTFYD--LAGNVMKNTFIBDSVGNMY 1149
Db 1041 YENVASDT-LFLFSSLLGKYYESIRV---DCKGYLYNSATFGDQVKASFITEA-GNLY 1094
QY 1150 FPDQGNKWNKHFVDVDSGEKGTGTFEFLKNGVFSFGGLVQTD--NGTYFDFNYGKMW 1207
Db 1095 YFGKDGVMYVTAQTIN-----GANTYFELNGTALR-NTIYTDAGQNSHYVANDGKRYEN 1147
```

QY 1208 Q 1208
Db 1148 E 1148

RESULT 8

JC5473

dextranucrase (EC 2.4.1.5) - Leuconostoc mesenteroides

C:Species: Leuconostoc mesenteroides

C:Date: 07-Jul-1997 #sequence_revision 29-Aug-1997 #text_change 29-Aug-1997

C:Accession: JC5473

R:Monchois, V.; Willemot, R.M.; Renaud-Simeon, M.; Croux, C.; Monsan, P.

Gene 182, 23-32, 1996

A:Title: Cloning and sequencing of a gene coding for a novel dextranucrase from *Leu*

A:Reference number: JC5473; MUID:97136686

A:Accession: JC5473

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-1290 <MON>

A:Cross-references: GR:U38181

C:Comment: This enzyme catalyzes the transfer of D-glucopyranosyl units from sucrose

C:Genetics:

A:Gene: dsrA

C:Keywords: glycosyltransferase; hexosyltransferase

F:78-870/Domain: catalytic #status predicted <CAT>

F:922-1290/Domain: glucan-binding #status predicted <GCB>

Query Match 33.9%; Score 2250; DB 2; Length 1290;

Best Local Similarity 46.4%; Pred. No. 2, 7e-107;

Matches 470; Conservative 155; Mismatches 336; Indels 52; Gaps 20;

QY 211 INGQYYIDPTTGQPKNFELLOGSNWYFDSDTGVTGNALBLOFAKGTSSNEQYRGN 270

Db 7 VDGKVFYF-DGQPKNFETIIDGKPYFYFDKDTGALSNN-DKQVYSELFSTGNKH---N 61

QY 271 AAYSYDDKSIENWVGLTADTYRPRKILKDGTTWDSKETDMPILMVMWPTLTQAYY 330

Db 62 AVYTSDDNFTQEGHLTASSWIRPRDILKNGKRAPSTVTDFFRLMAMWPKDSTQVY 121

QY 331 LNYMKOHGNLPSALPFFFNADAPAEALNHYSEIVQONIEKRISGTNGLTKLMDHFTV 390

Db 122 LNYMKDQ--LLSGTHFSDNENMRTLTAQAQAQVNIKKGIGLGNTGLTATQVID 179

QY 391 NRPMMNKDSNVNFSIQPGGFLAYENSIDLTPYANSYRLLGRMPINIKDQ---TYR-- 445

Db 180 AQPNNWIDSEAKGDD--HLQGGALLYTNSDMSPKANSYRKLSTPKQKQGIADKYQG 237

QY 446 GQFELLANDIDNSPNVQAEQLNWLNYLLNFGTITANNQANFDSVRVDAPDNIDALMN 505

Db 238 GFELLANDVDNSPNVQAEQLNWLHYMNIQSLQNDQANFQDGYRVDVNDVADLLQ 297

QY 506 IQADYFNAAVGHMDSDAVSNKHINILEDNNHADPE---YFNKIGNPQLTMDDTIKNSLHG 562

Db 298 IAGEYAKAAVGVDK--MTRERINIVQFKTGEMKIQTMSKHHATSKLSHDFPLHLAKYA 355

QY 563 LSDATN-RWGLDAIVHQSADRENSTENVIIPNYSFVRAHNNNSDQIQONAIR----- 616

Db 356 LNNPNDKRSGLPTREHSLVKRITDDKENVAQPNYSFIRAHDSVQTIITADIKKINPA 415

QY 617 VTGKDYHTTEFEQKGDIDAYIQDQNSTVKNLYNPASVAILLTKNDTIPRVYGYDLY 676

Db 416 STGLD-STVTILQIQAEFIYNADELKADKYVTPNIPASVAILLPNKDTPRVYGNF 474

QY 677 TDSGQYMEHQTRYDTLNLKSRVYKYGQSGQMTMSVGGNNNTLTSVRYGKGNATAD 736

Db 475 TDDQYMAKQSPYQDAIDALLKARIKAAAGGOTMKMNYFPDEQSVMTSVRYGKGNATAD 534

QY 737 TGTDETRTGIGVYVSNTPNLKLGVDKVLHMGAAHKNQYRAAAVLTDTGVNITYSDQ 796

Db 535 SGNQETRYOGIGLVNPNRPDLKLSOKDEYKMDMGAAHKNQYRPVLLTTKSGLYKYSTDA 594

QY 797 GAPVAMTDENGDLYLSSHNLVYNGKEEADTAVQGYANPDVSGVILAVVPGASDNQDART 856

```
Db 595 NAPVVRTDANGLT-----KAD-WYGVNDPQVSGYIAAWVPVGSANQDART 642
Qy 857 ABSTKNSGNSAYRNAAPDSNVIPAEAFNFTYPTKESERANVRQAQADPFASLGFTS 916
Db 643 KSETQSTGSGYHNSAALDSQVIYEGSNFODEFTPEFTNIIKIAQNVNLFKDWGLTS 702
Qy 917 FEMAPQYNSKORTFLDSTIDNGYAPTDRYDLGMSRPKNKYGTDEDLRNAIQALHAKAGLOV 976
Db 703 FEMAPQYRASDSKSELDATVQNGYAFTRDYICYNTPTKYGTADNLLDALRALHGGGQIA 762
Qy 977 MADWVPDQLYNLPGKEVAVTVTRDGRGNWKAALLNNLNVVNTTGGCEYKCKYGAFLD 1036
Db 763 INDWVPDQLYNLPGKEVAVTVTRDGRGNWKAALLNNLNVVNTTGGCEYKCKYGAFLD 822
Qy 1037 KIQKLYPEITFKKQVSTGVAIDESOKITEWSAKYFNGTNILHRGSGYVILKADGGYYNLG 1096
Db 823 QLKTOYQPLFOGKQISTDQPMFNDIQIKSWEAKYFNGSNIQRGAWYVLKDWGTQOYFNV 882
Qy 1097 TTTKQPLPQLHGEKKQGNQGVY-KGNDGNYFYDLAGNWNKNTIEDSVGNWYFDDQDG 1155
Db 883 SDAQTFLPKQLLGEK--AKTGVTYRKETS--FYSTSGYQAASAFICDN-GNWTYFDDKG 937
Qy 1156 KAVENKHFEVDVDSYGEKG--TYFELKNGVSGFGLQTDNGTYFEDNYGKAVRN 1207
Db 938 KAVVGNQVIN-----GANYTFLPNGLIELQAYLVHDDGMYIYINNIQGLHN 983

RESULT 9
A1483
Glucosyltransferase (EC 2.4.1.-) gtfs precursor - Streptococcus sobrinus
C:Species: Streptococcus sobrinus
C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 15-Oct-1999
C:Accession: A41483
R:Gilmore, K.S.; Russell, R.R.B.; Ferretti, J.J.
A:Title: Analysis of the Streptococcus downei gtfs gene, which specifies a glucosyltrans-
A:Reference number: A41483; NUID: 90316665
A:Accession: A41483
A:Molecule type: DNA
A:Residues: 1-1365 <GIL>
A:Cross-References: GB:M30943; NID:g153652; PIDN:AAA26898.1; PID:g153653
C:Genetics:
C:Gene: gtfs
C:Superfamily: cpl repeat homology
C:Keywords: glycosyltransferase; hexosyltransferase

Query Match 33.0%; Score 2191.5; DB 2; Length 1365;
Best Local Similarity 40.9%; Pred. No. 3e-104;
Matches 518; Conservative 170; Mismatches 436; Indels 141; Gaps 34:

Qy 51 AADGNSGGNTVTDQFSKNYATGNGNFYKYNQVPSGWHATNQSDKDSQWIIIVL 110
Db 36 AADTNNDGTSV--QVNK--WVPSPDKFDAQNG--QLAQAEPKAAQAQDATATSQVSPAT 91
Qy 111 NGEKVKROLVNDTKCEAGFNNDYKVPALNSMSGFCQIITLPTVTKENVQLVHR 170
Db 92 DGR-VDNQVTPAANQPAANVANDV--ANPATDAGAL-----NRQSAAD 132
Qy 171 FSNVKTGSGNTVDFWSESLPVPKDSQKNGPLKQFGQTINGQQOYIDPTTGQPRKNFL 230
Db 133 TSTDGA-----VP-QTSQDPGH-----LETVDGKYIYD-ANGQRLKNYS 171
Qy 231 LQSGNWTTFDSGTGVTGVALELQFAGKGVSSNEQVRNNAAYSDDKSIENVGYLRAD 290
Db 172 MVTDGRTYTFDGTGTEGQAQDLPRKTGAQNDVPDSYQANNQAYSNEASFVVDVNIAD 231
Qy 291 TWYRPRKILKDGTTWTDKSTDMRPTILMYWNTLTQAVTLANMKQGNLLPSALPFFNA 350
Db 232 SWYRPRKILKNGQWSQASSEGDLRPLIATWWDPAATKAAAYANFPAKEG-LISGSTRQNSA 290
Qy 351 DADPAELNHYSEIVQONIKERISETGNTDWTALHDFVTNNPMMNKDSENVNFSCIQ-- 408
```

```
Db 291 NLDAATQ-----IQSALEKKIASSEGTNWLDRKMSQFVKSONQWSIASNETVYPNQDH 345
Qy 409 FQGGFLKYENSOLDPYANSYRLLGRMP-----NIKQDYRGQETFLANDINDSNPVV 462
Db 346 MGGGALLFESKDETHANSRWLLNRNPTFQTKQKQFTTNYAGYELLANDVNDNSNPV 405
Qy 463 QAEQNLWYLLNFQGTITFANNQANFDSVRVADPNIDADLMTIAQDYFNAAYGMD-SDA 521
Db 406 QAEQNLHLHYLLNWGDIVAGDKANFDCGVRVADVNVNADLLQIQRDYVYKAKYGTQDNEK 465
Qy 522 VSNKHINILEDNHADPEYFNKIGNPOLTMDTITKNSLNHGLSDAT-NRWGLDAIVHQS 580
Db 466 NAIDLHLSLEAWSGNDNDYVKDQNNFSLSIDNDQRSGMLKAFGYASAYRGNLSLATAGL 525
Qy 581 ADRENNSTENVIPYSEVRADHNSQDQIQNAIRDVTGK----DYHTFTFEDBQKQIDA 636
Db 526 KNRSANPDSDPV-PNYVPIRADSEVQIRIAKIREKLGKTNADGLTNLLDLDLNAKAFDI 584
Qy 637 YIQDNSTVKYNLNLIPASAYILLTNKDTIPRVYVGLYTDGGQYMEHQYTRYDTLTNL 696
Db 585 YNQDMNATDKVYYPNLPMAWMLQNKDVTIRVYVGDYMDYTDNGQYMATKTFEYNAIETL 644
Qy 697 LKSRVYVAGQSM---QTMVGGNNNLTISVRYGKGAATATDTGTBETRCQICGVVYSN 753
Db 645 LKGRIKYYVAGQVSYKQDWSG---ILTSVRYGKGAASASDAGTETRNSGMALLINN 700
Qy 754 TPNLKLGVNDKVVLLHMGAAHKNQYRAAVLTATTDGIVNTSD---QGAPVAMTDESDLY 810
Db 701 RPNFRAYRN--LTLNMGAAHKSQAYRPLLLSTKDGIAIYLNDSVDVSRQYKYTDSQNL 758
Qy 811 LSSHNLVNGKEADTAVQGYANDPVSGYLAWVVPVAGSDNODARTAPSTKNSGNSAYR 870
Db 759 FSASEL-----QSVANAQVSGMIQVVPVGAADNQVRTSPSTQATKDGNIYH 806
Qy 871 TNAAPDSNVIPAEAFNFTYPTKESERANVRQAQADPFASLGFTSPMAQYNSKDRT 930
Db 807 QSDALDSQVIYEGFSNFQAFQSPDYTNVIAKNGDLFKSWGITQFEMAPQYVSSDGT 866
Qy 931 FLDSTIDNGYAFTRDYDLGMSPEPNKYGTDEDLRNAIQALHAKAGLOVMADVPDQIYNLPG 990
Db 867 FLDSVILNGYAFSDRYDLAMSNNKRYGSKQDLANAIGLQSAIGKVLSDLPNQLYNLPG 926
Qy 991 KEVATVTRVDDRGNVKWDALINNNLNVVNTTGGCEYKCKYGAFLDKLQKLYPEIFTKQ 1050
Db 927 KEVYATRVNYGQAKSGATINKTPTVANTIRSYGDYQBYGKGLDLDLQKLYPELFTKQ 986
Qy 1051 VSTGVAIDPSOKITEWSAKYFNGTNILHRGSGYVILKADGGYYNLGTTTKQFLPQLT-- 1108
Db 987 ISTGKPIDPSVKITWASAKYFNGSNIQRGAKYVL-SSENGYLL-ADCKLFLPTVLNNT 1044
Qy 1109 -GEKKQGNQVYKNGDNGYFYDLAGNVMVKNFTIBDSVGNWYFFDQDGKMYENKHFVDVD 1167
Db 1045 YGQPOVSANGFISKNGGIRHL-DKNGQVRKNRFEIS-GSNWYFDSQGMATGKTIGND 1102
Qy 1168 SY-----GEKTYFFLANG-----VSFRG----- 1186
Db 1103 TYLFMPNGKQLKEGVWYDGGK--AYYIDNGRTNWKGVFEFRVDGQDKHRYFNGDGTIAI 1161
Qy 1187 GLVQFDNGTYFYDNYGKMYRNQITINAGAMITLD-ENGLIKASVNSDAEYPTSTFDVKM 1245
Db 1162 GLVSLDNLTLFDAYGYQVKGQVTINGKSYTFDADQGDVQTDNANPA--PQSQAGWKL 1219
Qy 1246 LDONK 1250
Db 1220 LGDNQ 1224

RESULT 10
T30858
glucosyltransferase - Streptococcus salivarius
C:Species: Streptococcus salivarius
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
```

C:Accession: T30858
C:Simpson, C.L.; Giffard, P.M.; Jacques, N.A.
Infect. Immun. 63, 609-621, 1995
A:Title: Streptococcus salivarius ATCC 25975 possesses at least two genes coding for proteins that are essential for virulence
A:Reference number: 220909; MUID:95122197
A:Accession: T30858
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1577 <SIM>
A:Cross-references: EMBL:L35928; NID:G662380; PID:G662381; PIDN:AAC41413.1
C:Genetics:
A:Gene: gtfm

Query Match 32.8%; Score 2175.5; DB 2; Length 1577;
Best Local Similarity 40.4%; Pred. No. 2.4e-103;
Matches 518; Conservative 189; Mismatches 415; Indels 159; Gaps 38;

QY 19 SAVSGFDITIKLTN---DOYQALNGLOVLLRFSKAADGNPSGDNT-VTDQFSKNYATTG 74
DB 99 SAVSQAFAQAQTSFVASQEVAVSQTO-----SSGQETQTEQVSQQTST- 145
QY 75 GNFDYKYNQGVESFSGHATNQSDKDSOMIYLVN-GKEYKQLVNDTRKGAAGFRN 133
DB 146 -----QVAGQTSQAQSTFVTEQARPR-----VLTNAAPALATRAADSTIRINA--NRN 191
QY 134 DYKYNPAIENSMSGFGIITLPVVK-----NNWQLVHRFSNDVKYGEQNY 182
DB 192 ----TNIYITASGTTPNVTIITGPNIPKPNVTIISPNGTREPNVTITQPNPK----- 241
QY 183 VDFWSELMPVKDSF-QKNGPLK-----QFGLQTNGQQYIYIDPTTGPQPKNFL 230
DB 242 -----PVQSQSQPNKPVQPNQPSLDYKPVASNLKTIQCKYIYVE--NGVYKKAA 291
QY 231 IQSGNNWYIFDSDTGVTNALELQFAKQTVSSNEQYRNGNAYSYDDKSIENYNGYLFD 290
DB 292 IELDGLYVFD-ETGAMVDQSKPLRADAPNNSIYAVYVQAYDTSKSFELHDLNFDAD 350
QY 291 TWYRKQILKQDITWDSKETDMRILAVWPNITLQAYILNMQHGNLPLSALPFFNA 350
DB 351 SWYRKQILKQDKNWTAETKDYRELLMTWMPDKVTQVNYLNMYSQQG-----FGNKTYTT 406
QY 351 DADPAELNHSIVQNTIEKRISETGNTDMLTLMHDFVTNPNMKDSE--NVNFSGIQF 409
DB 407 DMNSYDLAAAEIVORGIEERIGRENTWLRQLMSDRIKTPQGNSESDNLLVKGDL 466
QY 410 QGGFLKYENSOLTPYANSYRLGLRPNIK-----DQYRGQEFLLANDIDNSNPV 462
DB 467 QGGALTLFMSATSHANSDFRLMNRPTNQGTGRKYHIDRSNGGYELLANDIDNSNPV 526
QY 463 QAEQLNLYILLNFGTITANNQANFDSVRVDAPONIDALMNIADQYFNAYGM--DSDA 521
DB 527 QAEQLNLYILLNFGTITANNQANFDSVRVDAPONIDALMNIADQYFNAYGM--DSDA 586
QY 522 VSKNHNILEDWHDAPYFNKIGNPQTMDDTIKNS--LNHGLSDATNRWGLDAIVHQSL 580
DB 587 NAIHLNILEAWSYNDHQYKDKTGAQLSIDNPLREILLTTLFLRSKNYRSGSLERYITNSL 646
QY 581 ADRENNSTENVIPYNSFVRAHDNNSQIQONAI--RDVTGK--DYITTFEDEKQIDAYI 638
DB 647 NNRSEQRHTPRDANYIFVRAHDESEQAVLANIISKQINPKPTDGTFTMDLQKAFETN 706
QY 639 QDQNSVYKYNLYNPASVAILTNKDTIPRYVYGLVTDGQYMEHQTRYVDTLTNLLK 698
DB 707 ADIAKADKATYQNIIPAAATMLTNKDSITRYVYGLVTDGQYMAEKSPYINADILLR 766
QY 699 SRVYVAGQSMQTSVSGNNNLTISVRYGKANTATDGTDETRTQIGVGVVSVNTPNLK 758
DB 767 ARKYVAGQDMKYTYKING-YEIMSRYGKAGEBANQLGTAETRNQGLVLTARPNMK 825
QY 759 LGVNDKVLHMGAAHKNQYRAAVITTTDGVINYTSDQAP---VAMTDENGDLYSSHN 815
DB 826 LGANDRLVYVYMGAAHKNQYRPLLLSKSTGLATYKLDSDVPAGLVRYTDMQGNLFTTADD 885

QY 816 LVYNGKSEADTAVOGYANPDVSGYLAVWVPVGAQNDQARTAPSTENKNSGNSAYRTNAAF 875
DB 886 -----IAGHSTVEVSGYLAVWVPVGAQNDQARTKASSTR-KGEQVPESSAAL 932
QY 876 DSNVIEEAFSN---FVYPTKESERANVRIAQNAADFFASLGFTSFENAPQYNSKORTFL 932
DB 933 DSQYIYEGFSNFQFVTPSYTNRV---IAQNAKLFKEWGITSEFAPQYVSSQDQTEL 989
QY 933 DSTIDNGYAFTRDYDLGMSPEKNKYGTDEDLNAIQALHAGLQVMAADWPDQIYNLPKKE 992
DB 990 DSIIEYVAFEDRYDIAMSKNNKYSLKDLMDALRALHAGISALADWPDQIYNLPKKE 1049
QY 993 VATVTRVDDRGVAKDAIINNLYVNT-IGGGYQKKYGAGFLDKLQKLYPETFTKKVQ 1051
DB 1050 VVTASRTNSYGTTPRPAEYINSLYAAKTRTFGNDFOGKYGAGFLDKLQKLYPETFTKKVQ 1109
QY 1052 STGVAIDPSQKITEWSAKYFNGTNIHRSGYVLKADG-GOYVNL--GTTTKOFLPTLT 1108
DB 1110 SNGRKLITNEKITORSKAYFNGSLQGTGARYVLQDNATNGYFSVKAQT---FLPKQMT 1166
QY 1109 GEKKQGNBGFVK-GNOGNYFYDLAGNNVNTFIEDSYGNMYFFDQDGKMYENKHFVDVD 1167
DB 1167 ---ETTSGFRVGDVQY--LSIGGYLAKNTFIQVGANQWYFEDKNGNMVYTGQVID-- 1219
QY 1168 SYGEGTYFFLKNKNGVSRGGL-----VQTDNGTY-----YFDNYG 1202
DB 1220 --GKK--YFFLDNGLQHRVLRQSGDGHVYIDPKVQAPNGFYDFAGPRQDVRFDGNG 1275
QY 1203 KVRNQTINAGAMITLIDENG 1223
DB 1276 QMYRGLHMDMGTTFYFDEKGTG 1296

RESULT 11
A44811
glucosyltransferase (EC 2.4.1.1) I - Streptococcus salivarius
C:Species: Streptococcus salivarius
C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 15-Oct-1999
R:Accession: A44811; S22726; S28809
R:Giffard, P.M.; Simpson, C.L.; Milward, C.P.; Jacques, N.A.
J. Gen. Microbiol. 137, 2577-2593, 1991
A:Title: Molecular characterization of a cluster of at least two glucosyltransferase
A:Reference number: A44811; MUID:92148377
A:Accession: A44811
A:Molecule type: DNA
A:Residues: 1-1518 <GTF>
A:Cross-references: EMBL:Z11873; NID:947526; PIDN:CAA77900.1; PID:947527
A:Note: sequence extracted from NCBI backbone (NCBI:81050, NCBI:81052)
C:Genetics:
A:Gene: gtf
C:Superfamily: cpl repeat homology
C:Keywords: glycosyltransferase; hexosyltransferase
F:1307-1326/Domain: cpl repeat homology <CP4>

Query Match 31.5%; Score 2093; DB 2; Length 1518;
Best Local Similarity 37.2%; Pred. No. 3.9e-99;
Matches 511; Conservative 197; Mismatches 414; Indels 250; Gaps 42;

QY 27 TIKLTNDQYQALNGLOVLLRFSKAADGNPSGDNTVTDQFSKNYATTGTFYVYKVNQ 86
DB 18 TIAYASVALAVIAGLSVT---TSSVSADETQDKTVTQSNSTGTTASLVTSPATK----- 69
QY 87 VEFSGWHATNSDKDSQMIIVLVNGKE---VKRLVNDYK---EAGAGFNNDVYKVP 140
DB 70 -----EADKTNTEAD---VLTAPKETNAVETATTTNQATAEATATADV--AVA 118
QY 141 AIENSMGFGQIITL-----PVTYKENVQLVHRSNDVKTEGNYVDFWS 187
DB 119 AVPNK-----EAVVTTDAPAVTTEKAEQATVKAQEV-----NTEVKAPEA 160
QY 188 ELMPVKDSFGKNGPLKQFGLQTNGQQYIYIDPTTGPQPKNFLQSGNNWYIFDSDTG 247

Db 2159 YFNNNGMTGVORINGYV--YCFNNDGAM-----NGFQS--INGNYLYNIYG 2204
 QY 1020 -----TTGGGYKKYGAFLDKLQKLYPEIF-----TKQVSTGVAIDPSQK- 1062
 Db 2205 CTYGTGQNTNSRNYFNSVSMVLTGAQNTIGYIFGDSNGIMLTGVTQTAGNTYDFSSNG 2264
 QY 1063 --ITEW-----SAKYFNGTNIHLHSGSYVLKADGGQYNIUGTTKQPLPQLGGERKQONE 1116
 Db 2265 TATGTWVILSNKTYFSPSLGKLTGFI-TVSGDNVYLDADGLVLOTGWTVVDGNKYILNS 2323
 QY 1117 -----GFYKNGDNGNYFVDLAGNMVKNFTFIEDSVGNWFFEDDGKWE------NKH 1163
 Db 2324 NGVRQGFGLTLNNKYF--DTNGVWQGFPTINNRYFNDGDMQTLWTIDYNNKY 2380
 QY 1164 VD-----VDSYGRGYFFFLKNGVSFRGGLVQTDNGTY-FDNYGKVRNQTIN 1211
 Db 2381 FDSGTGRLIGFQIID--GKK--YFNSNGIMLTG--VQTIGNKFGYFNGYVMTG---- 2430
 QY 1212 AGAMITYLDENGLIKASYSNDAEPTSTIDVGKMLDQ 1248
 Db 2431 ----VODIDNH-----TYNFNDGSGASTGWTKLSK 2457
 RESULT 14
 S32920
 cell wall-associated protein precursor wapA [similarity] - Bacillus subtilis
 C:Species: Bacillus subtilis
 C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 21-Jul-2000
 C:Accession: S32920; E69730; T47101
 R:Poster, S.J.
 Mol. Microbiol. 8, 299-310, 1993
 A:Title: Molecular analysis of three major wall-associated proteins of Bacillus subtilis
 A:Reference number: S32919; MUID:93302506
 A:Accession: S32920
 A:Molecule type: DNA
 A:Residues: 1-2334 <FOS>
 A:Cross-references: GB:105634; NID:9304177; PIDN:AA22883.1; PID:g304179
 R:Kunze, F.; Ogasawara, N.; Moser, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berte
 C:Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch
 A: Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
 Nature 390, 249-256, 1997
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Funa, S.; Galizzi, A.; Gall
 lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
 Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois
 A:Authors: lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maue
 Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetel
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon
 A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron
 akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
 T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
 A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
 A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
 A:Reference number: A69580; MUID:98044033
 A:Accession: E69730
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-2334 <KUN>
 A:Cross-references: GB:299124; GB:AL009126; NID:g2636442; PIDN:CAB15959.1; PID:g2636469
 R:Experimental source: strain 168
 R:Yoshida, K.; Sano, H.; Seki, S.; Oda, M.; Fujimura, M.; Fujita, Y.
 Microbiology 141, 337-343, 1995
 A:Title: Cloning and sequencing of a 29 kb region of the Bacillus subtilis genome contai
 A:Reference number: 224350; MUID:95219088
 A:Accession: T47101
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-2334 <YOS>
 A:Cross-references: EMBL:D31856; NID:9603765; PIDN:BA066656.1; PID:9603782
 A:Experimental source: strain BGSCL1
 C:Genetics:
 A:Gene: wapA; N17G

C:Superfamily: cell wall-associated protein wapA

Query Match 4.0%; Score 264; DB 2; Length 2334;
 Best Local Similarity 19.3%; Pred. No. 3.6e-05;
 Matches 316; Conservative 191; Mismatches 522; Indels 612; Gaps 86;

QY 1 MSLTARPDVAAYPSL--YNSAVSGDFTTK---ITNDQYQALNQQLQVLLRFSKAADGNP 56
 Db 315 LSVSDTTFMSAIPPTNTSASSQKWDANLKAIVLKTGYDKTGTGTNYAFMFNNL---KP 371
 QY 57 SDGNVTVDQFSKNYA-----TTGGNFDYVKVNGQVEFGSGWHATNQSHDK 101
 Db 372 IQNMTVTYKATLTKTYVAHSYGTGKATGLMLDTVNSNYONAKVTWNTKPKAS--KNICKADVH 429
 QY 102 DQWITIVLVNGKEVRQLVNDTKEGAACF-----NRNDVIKVNPAIENSSMSGFQGI 153
 Db 430 KGQWASYDVTA-----AVKSWNSGGANYGFKLHTNGKKEYWK---KLISSANSANKPY 480
 QY 154 ITLPVTYVKNENVQLVHRFSNDVKVGEHNYVDFWSELMPVKDSFQKNGPLKQFGLQTING 213
 Db 481 IEVYTIPIKGNITPTIKAYHNGDSTG--YFDI-----SWKKVEG--AKGYKVWIYNG 527
 QY 214 QQYY-----IDPTTGQ--PRKNFLQ----- 232
 Db 528 KEYQAIAGNVTWSFTGKKIWPMTSAETASKRYKULHDKDGAELADPSPVYKNSGGSY 587
 QY 233 --SGHNWTFDSDTGVGTNALEL-----QFAGTVSSNEQYRNGNAAYSVD----- 276
 Db 588 ATSKNYWIGVSAIFPOGSGAMSAPAKPIPVNUGKAQAPSAKGYNGNATGYFDLSWKAUS 647
 QY 277 -----DKSIENVGVLTADTYRPRKQILKDTGTTWDSKETDMRPIILMVWNTL---TQA 328
 Db 648 GATGYKVQVFNKGKGFETLD-----LGNQTSMT---TKGKKI---WPTSAREIRAGK 691
 QY 329 YVLNMYKQHGNNLP--SALPRFNADADPAELNHYSEIYOON----- 367
 Db 692 YALHLKDGSGAELPIMPPTYKAGGCGAKRNYSPKIIAYNKDGAIASPAATPALPDIA 751
 QY 368 -----TEKRISERG--NTDWLRT-----LMHDFVTNN 392
 Db 752 RPKWTVGYLYTNKSSQGTGVNVLWEKYQNAKGYKNVYNGKEVQSDVGDADHWTQNK 811
 QY 393 PMWNKDSNVNFGSQF-----QGGFLKYSNDSLTTPYANSYR-----L 431
 Db 812 NIW-PTSEIKAGSYKLTHTDCKGGELADPSPVTNNANGYKGGKNYSFTLVAYDANGET 870
 QY 432 LGRMPINIKDQTYRQEFLLAND-----IDNSNPVQASQLAWIYLLNFGTITANNOAN 487
 Db 871 IPTAPFN--PTFHEGAELFTEEYSIID-----IPSQLNGA-----TGNVIYNEEDLS 918
 QY 488 FDS-----VRYDAPDNI-----DADLNIADYFNAAYGMDSDAVSNKHIN- 528
 Db 919 IDGRPGGLSLSTYNSLSDSHLFGQGWYADAETSVISTD--GGAMYIDEATTHRTFKK 976
 QY 529 -----ILEDWNHADP-----EYFNKIGNPQLTWDDYTKNSLNHGLSDAT 567
 Db 977 ADGTYQPTGYVLELTETADQFILKTKDQTNAYFNKGGRL----- 1017
 QY 568 NRWGLDAIVHOSLADRENSTENVVIPNYSFVRAHDNNSQDQIQNAIRDVNGDYHHTFF 627
 Db 1018 -----QKVVDGHNAT-----VYTYNDKQQL--TATDASGRKL--TFTY 1053
 QY 628 ED-----BQKGDYD-----IQDQNSTYKKNLYN---IPASYAILLTNKT 666
 Db 1054 DENGHWTSITGPKNKKVITYSYENDLLKKVTDGTVTSDYDSEGRLVKQYSAHTEAKP 1113
 QY 667 IPRVYGLDYFDGGGYMEHQTR-----YYD-----TITNLKSRVKYVAG 706
 Db 1114 VFTEY-----QYSGHRLKAIKNAKKTIVYVSYADAKTKLLMTQPNGRKVQGYNEA 1164
 QY 707 GQSMQTSVGGNNNILLTSVRYKGGAMTATDGTDETRQGGIGVVVSVNTPNKLGVNDKVV 766


```
Db 1165 GNPQIVDDAEGKLTNTKY-----BGNNVVEDVDPN-DVGT----- 1201
Qy 767 LHGAAHKKQYRAAVLTDTGIVNTSOGAPVAMTDENGDLYLSSHNLVNGKEEAT 826
Db 1202 ---GKATESYQY-----DKDG--NVTSDKAYGTETE-----YKNNDVTKMKDTEG 1246
Qy 827 AVOGYANPDVSGYLAVWVPVSGASDNQDARTAPSTKXNSAVRT-NAAFDSNVI---F 881
Db 1247 TDIAYGDLA-----VSETDQSGKSSAAVYDKYGNQIQSSKDLASTNILKDGSGF 1297
Qy 882 EA-FSNFVYPTTKESERANVRIAQNAFDFAISLFTSFEMAPQYNSKDRFTDLDSTIDNGY 940
Db 1298 EAQKSGWNLTKADRRKISV-IADKSGVLS--GSKALEVLSQSTS-----AGTDHGY 1346
Qy 941 -AFTDRYDLGKSEPN-----KYGTD-----EDLRN---AIQALHK--AGLQW 977
Db 1347 SSATQIVEL---EPNTYTLTGKIKTDLAKSRAYFNIDLRDKQKRIQIHNHESALAGK 1403
Qy 978 ADWVPDQI-YNLP---GKEVAVV---TRVDDRGVYMKDAIINNLYVNTIGGGEVQK 1029
Db 1404 NDWTKRQITFTPANAGKAVVYMEVDHDKDKGKAWFDEV-----QLEKGEVSS 1454
Qy 1030 Y-----GGAFLDKLQKLYPEI-----FTKKQY----- 1051
Db 1455 YNPQNSFTSATENNVSASVDSBEGFNDDVSLKAARTSQAQSVTKQTWVLQSAN 1514
Qy 1052 -----STGVALDPSQKITE-----WSAKYNGNHLHRSYVYL 1085
Db 1515 DKPVYLTGMSKASSVKFTDEKDYSLQANVYADGSTGIYNAKFPSTQEWNRRAVVP 1574
Qy 1086 KADGGQYNLGTTKQPLQTLNGEKGQNEGVKNDGNVYFDL---AGMVKNTFI 1141
Db 1575 K-----TK---PI---NKVDLSILFQKATGVWFDDIRLEGSLTKSTY- 1614
Qy 1142 EDSVGNWYF-----FDQDGKVENKHFVDVDSYGEKGYFF-----L 1178
Db 1615 -DSNGVYVTKDELGATSTYDTERGKTS-----ETDAKGEKTYTYDQADQLTNMTL 1668
Qy 1179 KNGVSRGGIVQDNGTYFDYDNGYKVRNQTINAGA---MYTLDENGKLKASYNSD-- 1233
Db 1669 SNGTSI-----LHSDYKEGNEV-SKTIKAGADQYKFEYDVGKLVK---TTDPL 1714
Qy 1234 -----AEYPTSDVGRMLDON 1249
Db 1715 GNVLASEYDANSNLTKTISPN 1735

RESULT 15
D89824
hypotheical protein sdrD (imported) - Staphylococcus aureus (strain N315)
C:Species: Staphylococcus aureus
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C:Accession: D89824
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc
ma, A.; Mizutani-U, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A:Title: Whole genome sequencing of meticillin-resistant Staphylococcus aureus.
A:Reference number: A89758; MUID:21311952; PMID:11418146
A:Accession: D89824
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1385 <KUR>
A:Cross-references: GB:BA000018; PID:g13700454; PIDN:BA041751.1; GSPDB:GN00149
A:Experimental source: strain N315
C:Genetics:
A:Gene: sdrD

Query Match 3.8%; Score 251.5; DB 2; Length 1385;
Best Local Similarity 20.6%; Pred. No.7e-05;
Matches 240; Conservative 135; Mismatches 381; Indels 407; Gaps 61;
```

```
Qy 257 KGVSS--NEQYRNGNAAYSDDKSIENWNG-----YLTADTWYRQKQLAD-OTTWTD 307
Db 93 KEWVSQGGNETSNGK--SIEKESVOSTGKNGKVEYSTAKSDEQASPKSTNEDLTKQTI 150
Qy 308 SKETDWRPFLAWWPNLTQAVLYNVMQHGHLPSALPFPFNADA--DPAELNHYSEIVQ 365
Db 151 SNQSGLOPDL-----ENKSVVQVTNEENKAVDAKTESTTLNYSKDAIK 196
Qy 366 QNIEKRISSETGNTDMLRTLHDFVTNNPMNKNDSNVFSGIQFGGFLKYNSD--LRPY 424
Db 197 SNAE-----TLVDNNS--NSNNEN-----NADILPK 221
Qy 425 ANSDYELLGRMPI-----NLKDYRGQEFLLANDTNSNPVVOAQLNWLVL 473
Db 222 STAPKSLNTRMAAIQPNSTDSKNVND--LITSNTFLAVVDADNSKTIIVPAOD-----YL 275
Qy 474 LNFGTITANNQANFDSVRDAPDNIDADLMAIAODYFNAAYGMDSDAVSNKHINILSDW 533
Db 276 SLKSQITVDVKYKSGDYFTIKYSDVQ-----VYGLNPEDI--KNIGDIKDP 320
Qy 534 NHADPEYFNKIGNPQTMDDTTKNSLNHGLSDATNRW-----GLDAIVHQ----- 578
Db 321 NNGETIATAK-----HDTANNLITYFTDYVDRFNSVKMGINYSIYMDADTIPVDKK 372
Qy 579 -----SLADRENSTENVVINYSEVRAHDNNSQDQIONAIRDYTGKDYHTFTFED--- 629
Db 373 DVPFSVTIGNQITTTADITYPAY---KEADNNS--IGSAFTETVS---HVGNEVDGY 423
Qy 630 -----EQKG-----IDAY-----IQDQNSTVKKYLNLYNPASAYAILLTK-- 664
Db 424 YNQVYVNPMDKDLGAKLKVEAYHPKYPNTNIGQINQNVNINIKIYRPEGYTL---NKG 480
Qy 665 --DTIPRVYGDLYDGOYMEHQRYVD--TLTN---LLKSRVYVAGGQ---SMQT 712
Db 481 DYNTNDLVDTDEFKMKMTYGSQSNLDGDTISAVVAVVWNTKFKQYTNSESPTLVQMAT 540
Qy 713 MSVGGNNLTLSVR-----YKGKAMTATDT---GTDYETQIG---VV 750
Db 541 LSTGKNSVSTGNALGFTNNQSGGAGQEVYKIGNVTWEDTNKNGVQELGEGVGWNVTVV 600
Qy 751 VSTPNLKLG--VNDKVLHMGAAHKNQYRAAVLTDTGIVNTSOGAPVAMTDENG 808
Db 601 FDNNTNTKVEATKEDGSLIPNLPGDYRVERFNSLPKGYEVTTPSKQGNNEEL--DSNG- 658
Qy 809 LYLSSHLNVVNGKEE-----ADTAVQGYANPDVDSGYLAWVWPVGAS 849
Db 659 --LSS--VITVNGKDLNSADLGIYKPKYKILGDIYVWEDTNKNGIQDQDEKISGTV--- 710
Qy 850 DNQDANTAPSTKXNSAVRTNAAFDSNVIFEAFSNFVYPTTKESERANVRIAQNAOFF 909
Db 711 -----TLKDENGVNLKT-----VTTDADGKYKFT---DLDCNGVYK---EFT 746
Qy 910 ASLGFTSFEMAPQYNSKDRTEL-----DSTIDNGYAFTDRYDLG---MSEPNKY 956
Db 747 TPEGTPPTVTSQSDIEKDSNGLFTTGVINGADNNTLDSGYFKTKPKYKILGNVWEDTNKD 806
Qy 957 STDEDLNIAQLHAKGLQVMAWPDQIYINLPGRYAVVTVDDRGVWVKDAIINNLY 1016
Db 807 GKQDSTEKISGV-----TVTLKNGEYV-----LQ 832
Qy 1017 VVNTIGGGEYQ--KRYGGAFLDKLQKLYPEIFTKKQVSTGV--AIDPSQKITEMSAKYFN 1072
Db 833 TTKTKDKGKYQFTGLENGTY--KVEFETPSGYTTPQVSGTDEGIDS-----N 878
Qy 1073 GNNILHRSGSYVLK-----ADGGQY---YNLG-----TTTKQFLPIQLTGSK----- 1111
Db 879 GTS-----RTGVIKDKDNDTIDSGFYKPKYKILGDIYVWEDTNKN--GVQCKDEKISGTV 931
Qy 1112 --KQNEGEFVK---GNDGNYFYDLAGNWKNTF----- 1140
Db 932 TLKDENDKVLKTVITDENGKYQFTDLNNGTYKVEFETPSGYTTPSTVTSNDTEKDSNGIT 991
Qy 1141 ----TED-----SVGN--WYFFDQDGKVENKHFVDVDSYGEKGYTFFL 1178
```

Db 992 TTGVKADNMTLDSGKPKYSLGDIYVWYDSNKKQK-----DSTKGIKDVKIL 1044
Qy 1179 KNGVSFRGLVQTD-NGTYIFDN 1200
Db 1045 LNEKGEVIGTKTKDENGKIRFDN 1067
RESULT 16
AD1841
hypothetical protein alr0276 [imported] - Anabaena sp. (strain PCC 7120)
C:Species: Anabaena sp.
A:Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 11-Jan-2002
C:Accession: AD1841
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimizu, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena sp. PCC 7120
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AD1841
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2348 <KUR>
A:Cross-references: GB:BA000019; PIDN:BAE77800.1; PID:gl17135254; GSFDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr0276

Query Match 3.8%; Score 250; Db 2; Length 2348;
Best Local Similarity 18.8%; Pred. No. 0.00019;
Matches 285; Conservative 184; Mismatches 514; Indels 534; Gaps 72;

Qy 87 VEPGHWATWQNDKDSQWILVIVNGEVRQLVNDYKEAGFNRNDYKVPATENSS 146
Db 11 IALAGYNTTN---PDSIRIVLVV-----DIGAGTT-FNITD----- 42
Qy 147 MSGPGIITLPTVTKENVQLVHRFSNDVKTEG-----NYDFWSELMPVKD 194
Db 43 -NGHQ-----SSFRIGSILTYTAPQLISAGTVITWSSSNS 82
Qy 195 SFQKNGPLKQFGLQTNGQQYIDPTGPRKNFLGSGNNMIYFDSDT-----GVGTNA 250
Db 83 PGFNSNP-SNEALNASGDSLIITVGTGLASPTLIYALSSG-NWTNATSASTSAEPGANG 140
Qy 251 LELQFANGTVSSNEQYRNGNAAYSDDKSTENYGY---LTAATWPRKQLKDGTTWD 307
Db 141 GLETKTVAIT---TNG-----TNGYYSAGTAGTQVELLAASINPANWTA 184
Qy 308 SKE-TMRPILMVWPNLTCAVYL-----NYMKQHGILLPSALPFENADAPAEINHY 360
Db 185 SSTITDIAN---WPSFSLQPLPNIQITEYMYQGTN--GHEMETNLGTTAVDTGW 237
Qy 361 S-----EIVQMIKRISETGNTDMRLTLMHDPVTNNPWNKDSNVNF 404
Db 238 SYDDSGRTAGTSLSAFGLVQGESVILTEADEF-----RAAWGLS----- 280
Qy 405 SGIQFGGFEK-----YENSDLTPYANSYRLLRK-----PINIKDTYRG- 446
Db 281 ANKVIIGGLFRNIGRSEINLXN-----NNGQLIDRLTYGDETFPGTIRTRSGW 333
Qy 447 -----QPELLANDIDNSNPVQAEQLNWLXYLLN-----FGTITA 481
Db 334 TEPCLGAVTINTDQWLSTVDDQNSRTSGGIGNPCFNINNTPLPGITITQSGSTD 393
Qy 482 NNDQANFDSVRVADPNIDADL-MNIAQDYENAAAGMSDAVSNKHINIL--EDWNRHADP 538
Db 394 VTEGGVDSYTIIVLKTQPTADVITNIT-----VDNQVTTSSPLIFTPQNNIAQT 444
Qy 539 EYFNKIGNPQLTWDDTTIKNSLHGLSDATNHWGLDAIVHQSLADREN-----NTENNVVIP 594
Db 445 VTVAVVND--VIEGHTSTIAHNVSSTDTNNGYIAIANININITDNDAPPNTNVNQIT 502

Qy 595 NYSEVRAHDNNSODQIQNAIRDYTGKDYHTFTFEDBQK-----GI-----DAYIQD 640
Db 503 EYMTGANG-----EFVEFTNLGTTPTVDFTCWSFDDNTRIAGSENLSAFGLVQPGESVILT 558
Qy 641 QNSTVVKYHLYNIPASVAILTNKDDTIPRYYGDLTYDQGYMEHOTRYDYTLTNLKS- 699
Db 559 ETAAETFRANLPTSVKIIGNSNOGLGRADEINLSTGQLIDRLAYNDEVFTGVIITQ 618
Qy 700 -----RVKIVAGGQSMOTMSYG----- 716
Db 619 NASGWTAAHLDAFEITYNWQLSAINDQNSR-LSTGNDVGNPGTIPNPVSTVGAPKIT 677
Qy 717 -----GNNIILTSVRYGKAMTAT-DYGTDETRTQGGIGVVSNT-----PNLKLG 761
Db 678 VNPSTDFLDGQNLVPLPQIGAGALSIVINDPTPARTLGINFTLSDTDTVPVENLITIV 737
Qy 762 ---NDKVV-----LHMGAHKNQOYRAA-----VLTTDG-----VINYTSQOGAP 799
Db 738 TSSNQAVVPDANLTLTGTAERMLKINPAGVGLANTLTATVSDGTLSSSYIINYAASAGS- 796
Qy 800 VAMTDENGDLILSSHNLVNVNGKEADTAVQGYANPDVSGYLAVVWPVPGASDMDARTAPS 859
Db 797 -----VSPDRFELTGTSDASSAIALDAN-----YMFV-----ADDEDQTIIRLY 834
Qy 860 TEKNSGNSAYRTNAATD-----SNVIFRA-----FSNFVYTPKESERANVRYAQA 906
Db 835 DRNSGLPL-----ASFEDTSLGLSGSSEVDIEASTRIGNTIYWSGSHSNNSNGNDSPNR 890
Qy 907 DPE-----ASLGFTSFMAFQXNSSKDKTFLDSTI-----DNGYATDRYDLOM 950
Db 891 ERIPATQISGTASALITFQGYQ-----FLEDDLIAMDNNGHGLGAGF-IGL 938
Qy 951 SEPKNYGTDEDLNATQALHAKAGLOVMAD---WVPDQYINLPKGEVATVTRVDDRGVW 1006
Db 939 AESAANGVSGPEIRNGENI---EGLTVPDNTAYVSRAPNQ-----TSRDTN-- 984
Qy 1007 KDAIL---NNNLVYVNTIGGGEYQKKYGG-AFLD-----KLOKLYPEIFTKQV 1051
Db 985 --ALIIPVTNFTNLTNLTGTGSAASFGAPIFLDLGGVGRVSRSTERNSSNOYLITAGPPGG 1042
Qy 1052 STGVAID-----PSOKITEWSAKYENCT-----NILHRGSGYVLKADG 1089
Db 1043 STGIAPNDFRFLYTWGNADTAPVLRISDITALTNTNSPESIVEPDSLNTNTQIQLVDN 1102
Qy 1090 GQ--YINLGTGTTK-----QFLPIQLTGE 1110
Db 1103 GDTVYVNGTISKDLAQDNLQKPRSEVITLGPALISLTPYSGEENSLISSGS 1162
Qy 1111 KKGNGEFPVKG-----NCGNYFYDL-----AGNMVKNTF 1140
Db 1163 ETWTDSTIAGWTARTGTGTTIVASTGNTAGNLVSFGLDSSDRALSGISGNAAGTF 1222
Qy 1141 IEDSVGNWYFFDQDKHVENKHFDVDSYGEKGTVEFLKNGVS-----FRGLVQT 1191
Db 1223 YWGS-----REFNDTGNVNOLY---VNYIGEQ---WRSQGTTSNPQTVDYQYQATSL 1271
Qy 1192 DNGTYIFDN---YGKAVRQNTINAGAMITLDENGKLKIKASYN-----DA 1234
Db 1272 TGTGTWYDVANLDTSLINNT--AGALNGNANRNLISGTLISGLINPGEIWRWVDI 1329
Qy 1235 EYPTSDVCKMLDQNKI 1251
Db 1330 DHP-GTDHGLSIDDKV 1345

RESULT 17
A64556

toxin-like outer membrane protein HP0289 - Helicobacter pylori (strain 26695)
C:Species: Helicobacter pylori
C:Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 08-Oct-1999
C:Accession: A64556
R:Tomb, J.F.; White, O.; Kervatage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann,

Db 985 KDSNGLATTGVIKADNNTLSDGFYKTPKYSGLDYVVDNSKDGKQSDTEKGIKDVKVTLL 1044
 QY 1169 YGEGTYFFLANGVSFRGLVQTD-NGTYTFDN 1200
 Db 1045 LNERGEVI-----GTRKTDENGKICFDN 1067
 RESULT 20
 toxin-like outer membrane protein jhp0274 - Helicobacter pylori (strain J99)
 C:Species: Helicobacter pylori
 A:Variety: strain J99
 C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 08-Oct-1999
 C:Accession: C71953
 R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;
 Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;
 Nature 397, 176-180, 1999
 A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric path
 A:Reference number: A71800; MUID:99120557
 A:Accession: C71953
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-2902 <ARN>
 A:Cross-References: GB:AE001464; GB:AE001439; NID:94154789; PID:9415479
 A:Experimental source: strain J99
 C:Genetics:
 A:Gene: jhp0274
 Query Match 3.6%; Score 239; Db 2; Length 2902;
 Best Local Similarity 20.1%; Pred. No. 0.00095;
 Matches 294; Conservative 172; Mismatches 565; Indels 434; Gaps 76;
 QY 13 YPSLYNAVSGFDFTIKLTNDQYQALNGQVLRLFRKAADGNPSGDNVYDQPSKNYAT 72
 Db 366 HPQONSFSFGSATTLKGFATFEQAFNNS-NHQLTIQNASFNNAFNNTCKITTEKDAF 424
 QY 73 TGGNFDYVKNVNGNVEFSGWHATNQSDKDSQWILVNVKKEVRLQVLNTEKGA--GFN 131
 Db 425 NWTSEF-TPVDYTNMTISGG-----VTLSEK-----NDLKNGATLDGF 461
 QY 132 RNDVYKNPAIENSMGFGGIIPL-----PVTYKNEVQLVHRFSNDVKTGRCNYDFW 186
 Db 462 SSKITLQAGTFTNLSLSEKSVTLNRSRGITVHLLNHNALSLTNALATNE----- 514
 QY 187 SELMPVKDSQKQ-----NGPLKQGLQITNCQQYIDPTGQPKNFLLQSGNNMIY 239
 Db 515 SSSKP--QSFAGQLMDITTYGTQL-----LNENATSKPTDSSPSK-----STNSTQV 564
 QY 240 FDSDTGVTNLALELQFAKGVSSNEQYRNGNAAYSDDKSIEN-----VNGYLTA-D 291
 Db 565 YQVGYKIGDIYKIQ---ETFSHNSIIIQALESGTYTPPPVINGSKFDLSASWYNADMP 621
 QY 292 WYRPKQILKDGTTWDSKETDMPILMW-----NLTQAY--YLYNMKQGNLIPS----- 320
 Db 622 WYHKKYIPKSNQETES--GYIYLPVQVINGSYTNPKQTFESASNSNLVIGYNATWTDHNV 680
 QY 321 -----WP-----NLTQAY--YLYNMKQGNLIPS----- 343
 Db 681 SSDDTAFAGTSGSALNGHGGPPYQCTGTTNGTYSAYHYITANLRSNRIGTGGGAN 740
 QY 344 -----ALPFPNADAPAEINHYSEIV---QONIEKRISSETGNTDRLTLMHDFVTNNP 393
 Db 741 LIFNGVDSINATATVQHAGAYSSMTSTQNMDNSQNLGNSGKLLVYGTFTNQ 800
 QY 394 MNKDS-----ENVNFGS--IQFGGFLKYENSDLTTPYANSRLLGRMPINIK 440
 Db 801 A--KDGKFIENAGQATPENTNFGSGYQFGSDLSLNSNN--QENS-----GSFEIGAK 850
 QY 441 DQTYRGQEE-----LLANDIDNSNPVQ-----AEQLNWLYYLLNFGT 478
 Db 851 NTFPNANFNSTSFNPNNSATTSFVGDFTNANSNLQIAGNAVFNGSTNGSQNTANFN 910

QY 479 ITANN--DOANFDSVRVDAFDNID-----ADLMNIAODYFNA---AYGMDSDAVSNKH--IN 538
 Db 911 TGSYNIACNATEDNVVFNSTPTSVKGVKVTNNITLKNLNAPLSFG-DCTIVFSAISVIN 969
 QY 529 ILEDMNHADPEYFNKIGNPOLTWDDTTIKSLNHLGSLDATNRMGL-----DAIVHQSLADR 583
 Db 970 IGEAITN-----GNP-ITLVSSSK-AIEYNDAFSKNLWOLINVOGHGASSEKLVSS 1018
 QY 584 ENNSTENWVIPNYSFVRAHDNNSODIQ-----NAI---EDVTGKDYHTFTFEDEQKID 635
 Db 1019 AGNGYDGVV--YGF-----NNQTYNPQEVFSPNSISIRLGVGM-----VFDYDMKSDR 1067
 QY 636 AYIQDQNSHTVKKXNLYNIPASAYAILTNKDTIPRVYYGLDLYTD---GGQYMEHQTYRYDT 692
 Db 1068 LYQN-----ALGFMTY-MPNSYNNNLGNLN--TIYYDNSIDFYASGKTLFWKAPESQT 1120
 QY 693 LTNLLKSRVYVAGGSMQPMVSGNNNIIITSV-----RYG--KGAMTATDTGTD 740
 Db 1121 FT-----GQN--SAIVFGAKNIWTSVSDAPQSNVIIRFGDNKGA-GSNDASGH 1165
 QY 741 ETRTQIGIGVVVSNTPNKLGVNDKV-----VLHWGAHKKQYRAAVALTTDGVINVTSD 795
 Db 1166 CMWLQICIGITGHYEAQKIYITGSIESGNRISGGGASLNFNGLOGILLNATLYN--- 1221
 QY 796 QGAPVAMTDENGDLIYS-SHNLYVNGKEEADTAVQGYANPDVS-----GYL 840
 Db 1222 ---RAAGTQSSSMNFVSNNSANIOANSYFIDDTAQNKNPNFNFNALNLDFSNSSFY - 1277
 QY 841 AVWVPVGSAD-----NODARTAPSTEKNSNSAVRTNAAFDSNVIFEARSNFYVTTKES 895
 Db 1278 ---VGQTSQVFKFNVAISFTNSNLSGLYQMQA---KSVLFDN-SNLSVSVGTSS 1328
 QY 896 ERAN-VRIAQNADFFASLGTSTFEMAPQYN---SKDRTEFLDSTIDNGVAFDTRYDLGM 950
 Db 1329 IKANAINLSQNASINAS-NHSTLELOGLDLNLTSLNLQSAINSNNTINDYASLIA 1387
 QY 951 SEPKNVGTDEDLNATQALKAGLQVWADWVPQIYNLPCKEVATVTRVDDRGNNWKDAI 1010
 Db 1388 SN-----GSHLNFNGAVN-FNSANI-----TTSLSSSSIVFKGAV 1421
 QY 1011 INNKLKVVNTIGGGEQKYGGAFL--DKLOKLYPELFTKKQVSTGVADPSPKITEWSA 1068
 Db 1422 SLRGQFNLSNNSLDFQ---GSSAITSNTAFNFDYDNFASQSPITFHQALDVKVPLS--- 1474
 QY 1069 KYPNGNHLHRSGYVLKADCGQYNNLGTFTKQFLPOLTGERKKQNEGFKVGNDCNYYF 1128
 Db 1475 ---LGCNLLNPNNSVNLKNSQ-----LVFSDQGSNLNANIDLSLNGN--- 1517
 QY 1129 YDLAGNKKNTFTEDSVGNWYFFDQDGKAVENKHFVDVDSYGEKGYFFLKNQVSPRGL 1188
 Db 1518 ---KRVYNIQADMNGNMY-----ERINFF---GMRINDGI 1548
 QY 1189 VQPDNGTYTF---DMYGMK-----VRNOTINAGAMIYTLDE----- 1221
 Db 1549 YDAKNQTYSTNPLNNAALKITESFKNNQLSVTLQSOTPGIKNTLYNGSEIFNQKYVNA 1608
 QY 1222 NGKLIRASNSDAE---YFPTSDVG 1243
 Db 1609 NGVY---SYSDDAQGVFYLTSVKG 1630

RESULT 21

D71917

toxin-like outer membrane protein jhp0556 - Helicobacter pylori (strain J99)
 C:Species: Helicobacter pylori
 A:Variety: strain J99

C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 08-Oct-1999
 C:Accession: D71917

R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.
 Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.
 Nature 397, 176-180, 1999

A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric
 A:Reference number: A71800; MUID:99120557

A:Accession: D71917
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-3194 <ARN>
 A:Cross-references: GB:AE001488; GB:AE001439; NID:94155100; PID:AA06134.1; PID:94155100
 A:Experimental source: strain J99
 C:Genetics:
 A:Gene: jhp0556

Query Match 3.5%; Score 229.5; DB 2; Length 3194;
 Best Local Similarity 19.5%; Pred. No. 0.0034;
 Matches 284; Conservative 182; Mismatches 523; Indels 465; Gaps 75;

QY 39 NSOLOVLLRFSAADGNFSGDN-----TVTDQFSKNVATGGNFDVYKVGNGVQEF 89
 DB 5 NKKIDLERFSK-----NSRFGIKKIAKKYTIANSLSIYSLKTHSNSSLSINKKIFLG 60
 QY 90 SOWHA--TNQSHNDKDSQ--WITVLVNGKEVKRQLVNDPKGAAGFNNDVYKVP-----A 141
 DB 61 LGFVSALSAQSDYNSVYWL-----NSVNE-----NNKSYIISPLRTWA 102
 QY 142 IENSMGSGFGIITLPTVTKENYQVHRFSNDVKTGENTVDSELPVKDSF----- 196
 DB 103 GGNRSFTQNNYSQDIYGTNKAATPNH---SSWFEKGIGITGVFKRQDIYICAV 159
 QY 197 -----QKNGPLKQGLQ---TINGOQYIIDPTTGPQRKNFLQSGNNWTFPSDGTGVG 247
 DB 160 GSGNELKTGGGAILVFESSNELTNGAYF-----QNNRAGTQTSWLNLSNNSV- 208
 QY 248 TWALELQFAKTVSSNEQYRNGNAAYSVDKKSIEVNV-GYLTADTWYRPKILADGTTWT 306
 DB 209 -NLTNDF--GNQTPNGGFMVGRKITTNGSVNGNGFNDVDS-----NGAT-- 254
 QY 307 DSKETIDRPIVMPNLTQAYLYNMYKQHGKLNLPSPALPFENADADPAELNHYSEIVQ 366
 DB 255 -----TISGVTFNNCAUTY-----KGGNGIGGSIITNSN-----INHYKLNLA 295
 QY 367 NIEKRISGTNDWL-----RFLMHDFTVNNPMMKNDSENVNFSIQOGGFLAKYEN 418
 DB 296 N-----SVTFNNSTLGSMPNGNANTIGNAYILN-----ANNITFNNLTNGGFMVFN 343
 QY 419 SD-----LTPYANSOYRLGRMPIN-----IKQIYR--GOEFLANDIDN 457
 DB 344 SDAHVNFQTTTINNPTSEFVN---WPKVIINPNAIFNQNITPTIGNATLFS-MEN 398
 QY 458 SNPVQAEOLNLYLL-----NFGTITANDQANFDSV-----RVDAP 496
 DB 399 GN--IAYDDVNLNLIIRLKNQATKQNSKATSNNTHTYVTVYNGGLTYHFRQIFSP 456
 QY 497 DNT-----DADLMNIAODYFNAAYGMDSDAVSNKHINILEDNHAADPEYFNK 543
 DB 457 DSVLSQSVYVYGANLYTNSVNLHNVNFKLNINDRATTFYLNGLTWYTOARPAQT 516
 QY 544 IG--NPQIYMDITL-----KNSLMHGLSDATNRWGLDAIVHQSIL-ADR----- 583
 DB 517 YGKNSALVFENATTPWANGAIPKSNSTVRFGYEGVN-WGKTGYITGTFTADRVYITGM 575
 QY 584 -ENNSTENVIPNYSVRRAHNSQDIQNAIROYTGDXHTFTFEDESKGIDAVIQDN 642
 DB 576 MSGNGAQTGGGATLNFVGATEIN-----TAGATFNNLKT-----TSQN 613
 QY 643 STYKKNLYNIPASVAILLTKNDTIPRVYVGLTYTGG-----QYM 683
 DB 614 SYMTFALNGSGSGKINYSQSD-----FYD--WIDGGYDFGTGNGVDSVFNKRAYIKFQ 666
 QY 684 EHTRIYDITLWLLASRVKAVAGGQSMQTSV-----CGNNHLLSVRYGKGMAT-- 733
 DB 667 GAENSYFNKNTNPLAGNFKF--QKTTIEKSVLNDASAKFADGVNNAFNEKDFNGSGSFN 724
 QY 734 -AYDTGTDETRTGIGVYVSNTPNKLGVNDKVLHMGAAHKNQOYRAAVLTDTGVTNY 792
 DB 725 HACTNNAFNNSFGSGSFENAKQVDFNGNS---FNGGVFNFNTPKASFYNDFNVNNO 781

QY 793 TSDQGAQVAMTDENGDIY-----LSSHNL-----VYNGKEADTAVQYANPDVSGYLAV 842
 DB 782 FKINGAQDTFTSKGVYFMQGLLSLSVGYTYQLNKS----- 821
 QY 843 WYPVGAQNODA-----RTAPSTEKNNSNSAYRNAAFDSN--- 878
 DB 822 ---VGKDNHNNALYQMLRWTSGENPSGKLVDENKTAP-----NSAKIYNVQFTDNGLT 871
 QY 879 -VIFAFSN-----FVYPTKESERANYR-----IAQNADEFASLGFTSFEMA-- 920
 DB 872 YTIKENFNIGITLCTIGTYHCVNINDAFNLKWNWNNASNTVYILN-GMTTWKTAGT 930
 QY 921 ----POYNSKDRITDSTID--NGYAFTRDYDLGMSBPNKYGTDEDLRNAIQALHKA-- 972
 DB 931 GVFTDYSCTNSVLVFNQTPPLAGANTPSVSVGFSGKTS--GAEMGLVGYIOGVKXANQ 988
 QY 973 -----GLOMADWVPDQIYNLPK-EVATVTRVDDRG---NWKDALINN-NLYVV 1018
 DB 989 IDITGIRSGNAKTTGGGATLVFNAQERLNANLNNDXAGLQNSRMNFTVNNGLNMT 1048
 QY 1019 NTIGGGEYKKGAGFLDKLQKLYPEITFKKQVSTG--VAIDPSQ-----KITWSAK 1069
 DB 1049 N--ANFSNCTPHGGFNL---KANNTWQKSVSGSGNGFVDNANANGNAVKNVNSD- 1101
 QY 1070 YFNGTNILHRG---SGYVLKADGGY--YNLGTWTKQPLTOLTEKKQSGNEGVKNGDG 1124
 DB 1102 --NGTLIYKGGNSAGNSLTLENNTFNSYNINAKAQNLI-----FNNNSF---NSG 1147
 QY 1125 NYVYFDLAGNMVK--NTEIED-----SVGNWYFFDQDGKME----- 1159
 DB 1148 SYSFNDTKNTVKTGNTLINSPPFSLKGSVSDNNSFNIERDLTKTYTLTLLSGDNK 1207
 QY 1160 -----NKHFDVDSY-GEKGYVFFLANGVSFRAGLVQTDNGTYF---DNVCK 1203
 DB 1208 YNNQALADWVSKNLMDLIHYDGEQGT-----LLRTDNTVYFVQFTQSGNQK 1254
 QY 1204 MVENQPINAGMILY 1217
 DB 1255 FVEETFPNGSITY 1260

RESULT 22

I40884

Cytotoxin L Clostridium sordellii

C:Species: Clostridium sordellii

C:Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 15-Oct-1999

C:Accession: I40884

R:Green, G.A.; Schue, V.; Montell, H.

Gene 161, 57-61, 1995

A:Title: Cloning and characterization of the cytotoxin L-encoding gene of Clostridium

A:Reference number: I40884; MID:95369733

A:Accession: I40884

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-2364 <RES>

A:Cross-references: EMBL:X82638; NID:g1000694; PID:CAA57959.1; PID:g1000695

C:Superfamily: cpl repeat homology

C:Keywords: cytotoxin

Query Match 3.4%; Score 225.5; DB 2; Length 2364;
 Best Local Similarity 19.1%; Pred. No. 0.0034;
 Matches 254; Conservative 178; Mismatches 470; Indels 429; Gaps 67;

QY 25 DTWIKLTNDYQALNQLOVILLRFSKAADGNPSG---DNTVTDQFSKN-----YATTGQNF 77
 DB 1272 DALITKLKPYEDTNNRNL-----DGNTRSFIVPVITTEQIRKMLSYSFYGGSGSY 1323
 QY 78 DY-----VKVNGVQNFSSGHATNOSNDKSDWITVLVNGKEV-KROLVMDTKGAAAGF 130
 DB 1324 SLSLSPNNMIDLNUVENDTWI-----DVDNVVKNITTESDEIQKGLIENI-----LSKL 1375

QY 131 NRNDVYKVPAINSSGGGQIITLTPVYKVENVQLVHRFSNDVKTGEGYVDFWSEL 190
DB 1376 NIED-----NKILANHTINFG-----DINESNREISLTFIILED-----INIIEE 1419
QY 191 PVKDSFGKNGPLKQFGLQTINGQQYVIDPTGPRKFLQSSNWIYFDSGTG-----245
DB 1420 LVSKYK-----ILLSCMKLLENSDSIQOKI 1447
QY 246 --VGTNALELQFAKGTSSNQYRNGNAAYSDKSIENYNGYLTDATWYRQKILKDG 303
DB 1448 DHGFGNGHQYIPYVIDNETKNGFIDYSKKE-----GLTAEF-----1488
QY 304 TWTDSKETDMPILMVWPNLTQAYLYNMYKQHGNNLPSALPPFNADAPALNHYSEI 363
DB 1489 -----SNESIIINI-----YMPDSNNL-----FIYSKDLKDI-----1516
QY 364 VQONLEKISGTGNTDML--RTLMHDFVTNPNMKDSSENVFSGIQFG-----GGFLKYE 417
DB 1517 -----RIKNGDVKLLIGNYFKDDMKVSLFTEDTNTIKLGVLDENGVAQILKPM 1569
QY 418 NSDLTPYANSYRLLGRMPINIKDQTYRGQ-----EPLAND--IDNSNPVVAQOL----467
DB 1570 NNAKSALNTSLANFLESINIKNIFYNLDPNIEFLDTNFISSNSISGQFELICDKD 1629
QY 468 --NWLXYLLNFG-----TITANDQANF--DSVRVADPNIDADLMIAQDYFNAAYGMD 518
DB 1630 KNIQYFFNFKIKETSYLYVGNQNLIVERYSYHLDDSGNISSTVINFSQKYL--XGI- 1685
QY 519 SDVSNKHII--NILEDNHADPEYFNKIGNPQLTMDYT--IKNSLHGLSDATNRWGLD 573
DB 1686 --DRYVNVIIAPNLYATDEINTPYKKNYICFEVILIADANYINEKINVINIDLISRYVMD 1744
QY 574 AIVHQSLADRENNSTENVVPIYSFVRAHDNNSQDIQ--NAIRDVAGKDYHTTETFEQ 631
DB 1745 -----NDGSDILIAN-----SEEDNQPVKIRFVNVFASDFAAKLSNFSDQK 1789
QY 632 KGIDAYIQOQNSTVKKYNYLNPASYALLTNKDTIPRVYDGLYTD--GGQYMSEQTR 688
DB 1790 -----DVSVKIISTFSLAA-----YSGDFDYEFGLVSLDNDYF 1824
QY 689 YDITLNLKSRVYVAGGQSMQYVSGGNRI--LTSVRYGKGMATATDGTDETRTQ 746
DB 1825 YNSGNGWVGLI-YI--NDSLYTKPKPNLLITGFTIDGNKYFFDPTKSG-----AAS 1876
QY 747 IGWVYNSPTNKLGVNDXVLIHMAHKNQYRAAVLTITDGVINYTSQCAPVAMTDEN 806
DB 1877 IGEITIDCKDYF--NKQIILQV-----VINTSGLAYF-----APAGTLDEN 1918
QY 807 GDLYLSSNLVAVNGKEADTAVQGYANDVSGYLAWVPVGCASQNDARAPSTEK-----862
DB 1919 -----LEGESYVNFGLKINDGKIYFED-----NYRAA--VENKLLDDETYFNPKTGEALKG 1969
QY 863 --NSGNSAYRTNAFDSNVIFEA-----FSN-----FVYTPKE 894
DB 1970 LHQIGDNKY-----YFDNGINGTGFITINDKAVFENDGVNGVQYIEVNGKYFFYGRNGE 2025
QY 895 SERAVRIAQADFFA-----SIGFTSFEMAPQYNSKDRT-----FLD-----STID 937
DB 2026 ROLGVNTPDGFKFPKDDDLGTEGSELT--LYNGILNFGNKIYFFDISNAVYVWGCTLD 2084
QY 938 NG--YAFTRDYD--LGHSEPN--KYGTDRDLRQAIALHKAQGLQVMADWVPDQIYNLP 990
DB 2085 DGSYTFDDNRAEACIGLVINDCKYFDD--NGI-----2117
QY 991 KEVAVTRVDRGNVWKDAIINNLYVYVNTIGGGEYKQYGGAFDLKQKLYPFIPTKKQ 1050
DB 2118 ROLGFTI-----INDNIFYESG-----KIELGYQNI-----2145
QY 1051 VGTGVAIDPSQKITWSAKYFNGTILHRGSGYL-----KADGGQYINLGTITTKQLPI 1105
DB 2146 -----NGNYFYDESGLVLIGVFDPTDGKY-----FAPL 2175
QY 1106 QLTGKCKQGN-----EGFYKNGDNGYFYFDLAGNMVKNFTIEDSVGNWYFFPDQDKMYENK 1161

DB 2176 MIVNDNITGQAVKYSGLVNEDVDYFGET--YKTIETGMIENETDKYFYDPTETKAYNGI 2233
QY 1162 HFVVDVDSYGEKGTFFELKNGVSFRGLVQTONGTYTFYFNVGKAVRNQOTINAGAMLYTILDE 1221
DB 2234 NVVD-----DKIYFDENGIR-MRTGLISFENNWNVYFNEDEGKA-QGYLNKDKMFWFGK 2285
QY 1222 NCKLIKASVNS 1232
DB 2286 DGKMQIGVNT 2296
RESULT 23
D71612
Hypothetical protein PFB0540w - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
C:Accession: D71612
R:Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, A.; Fittie, Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A:Reference number: A71600; MUID:99021743
A:Accession: D71612
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1844 <GAR>
A:Cross-references: GB:AE001401; GB:AE001362; NID:g3845209; PIDN:AAC71897.1; PID:g38
A:Experimental source: clone 3D7
C:Genetics:
A:Gene: PFB0540w
Query Match 3.4%; Score 222.5; DB 2; Length 1844;
Best Local Similarity 17.9%; Pred. No. G.0033;
Matches 223; Conservative 182; Mismatches 501; Indels 339; Gaps 54;
QY 77 FDYKVGNGQVFSGWH-A-TNOSNDKDSQIIVLVNGK-EVKARQLVNDTKGGAAGFNND 134
DB 10 FDEYSLNN-----SNLNAVYVMSNEAHNF---LYNRNDAMYEYNSINGHMSNINNN 62
QY 135 VYKVPAINSSSGEQGITLTPVYKVENVQLVHRFSNDVKTGEGYVDFWSELMPVK 194
DB 63 TNNLQDAYINKELH-YMNSDKXINAKHONVHTATYNNMDKXNANNIILNHNHMTD 121
QY 195 SFQKNGPLKQFGLQTINGQQYVIDPTGPRKFLQSG-----NWIYFDSGTGV 246
DB 122 D-----QNYFYSTTKANKNFKENFNNNNNTVNSFYHTDN-- 161
QY 247 GTNALELQFAKGTSSNEQ--YNGNRAYSYD-----DKSIENVGYLTADTWYRQKILK 300
DB 162 -----NFIETFRNQEQDYVNNIINNENNNQVYKNNNNNN-----200
QY 301 DGTWTDKETDMPILMVWPNLTQAYLYNMYKQHGNNLPSALPPFNADAPALNHY 360
DB 201 -----PKNVVE-----SINKFNHIVNMQNFHF 223
QY 361 SETVQO-----NIEKRISGTGNTDMLTTL--MHDFTNPNMKDSSENVFSGIQFGGFL 414
DB 224 IPNISGKNGLNENKASLSHNVSVYSTISELHNF---NYVMNMDLNVDNMMNNNNNNM 280
QY 415 KYBNSDITPYANSYRLLGRMPINIKDQTYRGQEFLLANDIONSNPVVAQOLNWLXYLL 474
DB 281 NNN 324
QY 475 NFGTITANN--DQANFOSVRVADPNITD--ADLMNTAQDYFNAAYGKDS--DAVSNNHINI 529
DB 325 NKNMDVNNMMNNMMDVNNMMNNMMNNMMNNMMNNMMNNMMNNMMNNMMNNMMNNMMNN 382
QY 530 LEDNNHADPEYFNKIGNPQLTMDDTTKNSLHGLSDATNRWGLDAIVHQSLL-----580
DB 383 ITHFNNGISHNVNNSN--FSNNAHLDSNNMNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 440

QY 1019 NTIG---GGYOKKYGAFDLKLOKLYPELTKKQVSTGVAIDPS--QKITSARKYENG 1073
 Db 2152 DNGIVQIGVFDTSKGKYAPANTVNDNIY-----GQAVEYSGLVKRGV-DVYFGE 2203
 QY 1074 TNLHRSYVYLKADGGQYVNLGTTKQFPIQLTGKKGQNGEGFVGKNGDFTYDLAG 1133
 Db 2204 TYTETGWIYDMENESKYY-----FNP-----ETKRACKGINLIDIKYFF-DEKG 2249
 QY 1134 NMVNTFTEDSVGKWFYFDDGKM-----VENKHFVDVDSYGEKGYFFLKNVGSFRGG 1187
 Db 2250 --IMRTGLISFENNRYTFNNGEMQGYINIEDKMFY-----FGEDGV-----MQIG 2294
 QY 1198 LVORTDNGTYFDNGKAVRN---QTINAGAMITY-----LDENGLIKASVNSDAEYPTST 1240
 Db 2295 VENTPDGFKYFAHQNTLDENFEGESIN-----YTGWLDLDE-----KRYFTD-EXIAAT 2343

RESULT 25
 C48399
 ABC-type transport protein ydbA.2 - Escherichia coli
 C:Species: Escherichia coli
 C:Date: 19-Nov-1993 #sequence_revision 16-Oct-1998 #text_change 08-Oct-1999
 C:Accession: C48399; D64891; H64891
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Cohen, A.; Rose, D.J.; Mau, B.; Shao, Y.
 Biochimie 73, 1361-1374, 1991
 A:Title: Multiple IS insertion sequences near the replication terminus in Escherichia coli K-12.
 A:Reference number: A48399; MUID:92190338
 A:Accession: C48399
 A:Molecule type: DNA
 A:Residues: 464-2020 <MOS>
 A:Cross-references: GB:D85081; NID:g3041754
 A:Experimental source: strain K-12
 A:Note: sequence inconsistent with the nucleotide translation
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Cohen, A.; Rose, D.J.; Mau, B.; Shao, Y.
 Science 277, 1453-1462, 1997
 A:Title: The complete genome sequence of Escherichia coli K-12.
 A:Reference number: A64720; MUID:97426617
 A:Accession: D64891
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-839, 'LDLPLYFQTSVIT' <BLA1>
 A:Cross-references: GB:A600237; GB:U00096; NID:g1787665; PID:AAC74483.1; PID:g1787668;
 A:Experimental source: strain K-12, substrain M61655
 A:Accession: H64891
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 'M', 915-2020 <BLA2>
 A:Cross-references: GB:A600237; GB:U00096; NID:g1787665; PID:AAC74487.1; PID:g1787672;
 A:Experimental source: strain K-12, substrain M61655
 C:Genetics:
 A:Gene: ydbA_2
 A:Start codon: GTG

Query Match 3.3%; Score 221.5; DB 2: Length 2020;
 Best Local Similarity 19.6%; Pred. No. 0.0043;
 Matches 280; Conservative 168; Mismatches 513; Indels 465; Gaps 72;

QY 55 NPSGNTVTD-----QFSKNVAT-----TG-----GNEDYKVNQGVFSG 91
 Db 497 NNEGESTITNGTGQINGNDATANNKSTVDGKDSGTGKIAGNIGVINDGSLTDTGG 556
 QY 92 WHATQNSD-----KDSQWITVYNGKEVKQLVNDYKEGAAGFNNDYKVN 139
 Db 557 ABGVENIGDNGVNNKGDIVVDSDIGSIGVING-----EGATVNTGDVNVSN 604
 QY 140 PA-----IENSMSGFQGITL-----PVTVKNNVQLVHRFSNDVK-TGE 179
 Db 605 EATGFSITNSKVSLAGSMQGVDFSTGVDLNNGNNSVTLAAKDLKVVQKATGINVSGD 664
 QY 180 GNYVDVSELMVPKDSFKGNGPLKQFGLQTINGQOYYIDPTTG---QPRKNFLQSGNN 236

Db 665 ANTVNITGVLVDKDK-----TADNAAEFFDPSVGINVYVGSNNVTLQGL 711
 QY 237 WYIFSDP-----GVG-----TNALEQFAKGTSSNQYRNGN 270
 Db 712 TVVSDSEVTSRQSNLFDGSAEKTSGLVIGDGTNNMNGGLELIGKNALADGSGVTSLR 771
 QY 271 AAYSDDKXSIEVNG-----YLTADTWYRPKQLKDGTTWTDSKETDM-----RPI 316
 Db 772 TCYSY--TSVIVSGESSVYLNQD-----TISGEFPLGAGVIRVQDKAL 815
 QY 317 LMWNPNTLTQYVYLYNWKQGNLLPSLPFFNADADPAELNHTSEIVQCKIEKRISGT 376
 Db 816 LEIGSGATLM-QDIDSEFHHGTRP-----ELFYADSGAKIYNKG 855
 QY 377 NTDWLRTLMHDDVTNNMWNKDSNNVNSGIG-FOGG-----FLKYENSOLTPYAN 426
 Db 856 TVE-IQNLGFAEVT-----GENTTCINSISILLONGKDPAPSPVILLATNGSGATNACT 909
 QY 427 SDYRLGRMPINIKQTYRGQEFLLANDINSNVVQAEQLNMLYLLNFGTITANNDOA 486
 Db 910 ITCKVTEQHSVENKYSTGTSTNSFIENNDY-----SSITGLVAQS 948
 QY 487 NFDSDVRVDAFONLADLMLNIAQDYENAAVGM-----DSDAVSNKHINILEDNHADPEYEN 542
 Db 949 NSYII-----WTDGIDDL-----YGRSGVGLAIADSPAENQCKITIDSMVDA----- 993
 QY 543 KIGNFQLTMDDTIKNS-LNHGL-----SDATNRWGLDAIVHOSLADRENN 586
 Db 994 --NUTAMROIASNSAIDFGVGVGVSYSACKNATAINOLGGVITVYNAGAGMAAY 1050
 QY 587 STENVVPIVSVFRAHDNNSQOIQNAIEDYTKYHYTFEDROKGI-----DAYI 638
 Db 1051 GASNTVI-NQGTINLEKNYDSDLAANTLVGMAYEHTAINDQGTVININVGQAFY 1109
 QY 639 QQONSTVKY-----NLKNIPASY-AILTNKNDITPR----- 669
 Db 1110 NDCGTGVNKGITCFGVGCQSGNEYNNTDDFSLIYTGDDITRSGEITVNLKSAVTDK 1169
 QY 670 -----VYIGDLYTDGGQYNEHQTRYDYDLTLNLLSRVYVAGQSGMQMVSQVGN 720
 Db 1170 LAGNVVNSGTLSDQQTIVSSGULENTSG--GIINNLVLDKGAIVKNAGVNTNNVDVSG 1227
 QY 721 IITSVRYGKAMATATDT-----GUDERTQIGVYVSVNTENLKLGVNDKVLVHGAHAKNQ 777
 Db 1228 ILNNA-----GEMTAQITMAGADSS-----LVNNT-----GTINKIVQAGVFNNGS 1271
 QY 778 YRAAVLTITDGVINTYSDOGAPVAMTDENGDLVLSHNLVWNGKEADTAYQGYANPDVS 837
 Db 1272 VTCRMM-SAGGVFNQTD-CA-----IMRGAALTGTAV---ANNE--- 1306
 QY 838 GYLAVVVPVPGASDNDQARTAPSTEKNNSGNSAVTNAFDSNVIFEAFSNFVYTPKESER 897
 Db 1307 ---GTW-NLGSS-----SEGNTGMLEVNNNSAFNNR-----GEFID---DNK 1343
 QY 898 ANYRIAQNAADF-----ASLGPTSFEMAPQYNSKDKRTFLDSTID-----NG 939
 Db 1344 NAVHINQSTLYNTGRMINSNSSHNGAVNMWNGGNGRFINDTGIDVSAKSLVYSANNAGDQ 1403
 QY 940 YAPTDRYDLGMSFENKYGTDLDLNAIQALHAKAGLQVADVVP--DQIYNLPGKEVATVT 997
 Db 1404 NAFFWQDNGV-----INFDDHDSASAVK VTH-----SNFIAQNDGIMNISG--TGAVA 1449
 QY 998 RVDNRGNVVKDAIINNLLVWNKTIIGGGEYQKKYGAFLDKLQKLYPEITFKKQVSTGVAI 1057
 Db 1450 MEGDK-----NAQLVNNGTINLGTAG-----TTDTGMIGQL 1481
 QY 1058 DPOKLTIRSAKYFNGTNTILHRSGSYVLKADG--QGYVNLGTTTKQFLPIQJLAGEK--KQ 1113
 Db 1482 DAN--ATADAVIENGTINIFANDSEAFSVLGTGVGVHVNNGTV--IADGVTSGLLIQ 1536
 QY 1114 GN---EGFVKGDGN-----YFFDLAG-----NMVKNFTIEDSVGNWYFFDQCKAYE- 1159

Db 1537 GDSINVEG-MNCHNNGNSSEVHYGDTLPDPKPKNTVSVTSGSDEAG-----GSMNNL 1587
Qy 1160 NKHFVDVDSYGEKGYFFLK-----NGVSFRGLVQ--TDNGTYTFDNY---GKAVRNQT 1209
Db 1588 NGYVVGTVNNGSAGK---LKVNNAAMNGVEINTGFTAGTADTTVSFDNYVEGSLTDADA 1644
Qy 1210 INAGMIYT---LDENK---LIRASYNSDAEXPTSTDVGKMLD 1247
Db 1645 ITSTSVVWTAKGSTDASGNVDYTWSKNAYTDVATDASVNDIAKALD 1690

Search completed: August 12, 2002, 07:54:30
Job time: 107 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 12, 2002, 07:54:33 ; Search time 18.06 Seconds

(without alignments)
2682.068 Million cell updates/sec

Title: US-09-995-749a-2_COPY_531_1781

Perfect score: 6641

Sequence: 1 MSLTARPDVAAYVPSLYNSA.....SDAEYPTSTDVGRMLDQNK 1251

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : SWISSPROT_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2446	36.8	1462	1	GTFD_STRMU
2	2347.5	35.3	1375	1	GTFD_STRMU
3	2330.5	35.1	1476	1	GTFB_STRMU
4	2274	34.2	1592	1	GTFD_STRDO
5	2271	34.2	1597	1	GTFI_STRDO
6	2191.5	33.0	1365	1	GTFB_STRDO
7	264	4.0	2334	1	WAPA_BACSD
8	222	3.3	2366	1	TOXB_CLODI
9	221.5	3.3	1861	1	APU_THERU
10	216.5	3.3	1279	1	APU_THESA
11	213	3.2	828	1	PMEC_PROMI
12	210	3.2	2003	1	YDBA_ECOLI
13	209	3.1	2710	1	YDBA_ECOLI
14	206	3.1	1296	1	VACI_HELPY
15	204.5	3.1	1475	1	APU_THERY
16	201	3.0	1902	1	P2P_LACPA
17	199.5	3.0	1654	1	OMP8_RICRI
18	199	3.0	1902	1	P2P_LACLC
19	198.5	3.0	1643	1	OMP8_RICPR
20	198	3.0	1645	1	OMP8_RICRY
21	195	2.9	1300	1	120K_RICRI
22	195	2.9	1541	1	IGA1_HAEIN
23	195	2.9	1902	1	P1P_LACLC
24	191	2.9	918	1	YAMB_CAEEL
25	190.5	2.9	1694	1	IGA0_HAEIN
26	190	2.9	1287	1	VAC2_HELPY
27	189	2.8	1532	1	IGA_NEIGO
28	187.5	2.8	1104	1	COIA_CLOPE
29	187.5	2.8	1569	1	YPTA_ECOLI
30	186.5	2.8	1953	1	BIGA_SALTY
31	186	2.8	1288	1	VACA_HELPY
32	185	2.8	1655	1	OMP8_RICCN
33	183.5	2.8	1481	1	APU_THERY

34	183.5	2.8	1902	1	P3P_LACLC
35	181.5	2.7	1702	1	IGA2_HAEIN
36	181	2.7	1183	1	CNA_STAND
37	178	2.7	1290	1	VACA_HELPY
38	177.5	2.7	512	1	AMY_BACLI
39	177	2.7	1162	1	EXEN_CLOBO
40	176	2.7	1577	1	HYNA_PROMI
41	176	2.7	1157	1	HYNA_THESA
42	176	2.7	1162	1	EXEN_CLOBU
43	175.5	2.6	1656	1	OMP8_RICJA
44	174.5	2.6	1545	1	IGA3_HAEIN
45	174	2.6	1250	1	BXE_CLOBO
46	173.5	2.6	917	1	HXA3_HAEIN
47	172.5	2.6	1628	1	NAGH_CLOPE
48	172.5	2.6	1829	1	FRPC_NEIMB
49	172	2.6	1004	1	SLPO_BACBR
50	172	2.6	1770	1	YCB9_YEAST
51	171	2.6	1251	1	RBP2_PLAVB
52	170.5	2.6	1341	1	VG37_BPT2
53	170	2.6	1063	1	HGPI_HAEIN
54	169.5	2.6	1274	1	BXF_CLOBO
55	169	2.5	1026	1	VG37_BPT4
56	168.5	2.5	1296	1	BXG_CLOBO
57	168	2.5	1291	1	VAC4_HELPY
58	167.5	2.5	1385	1	CSAA_BACUD
59	164.5	2.5	549	1	AMY_BACST
60	164.5	2.5	1310	1	VAC3_HELPY
61	164.5	2.5	1829	1	FRPC_NEIMC
62	163	2.5	718	1	CDGT_BACCI
63	162.5	2.4	518	1	AMT6_BACS7
64	162.5	2.4	1169	1	C9DA_BACTP
65	161.5	2.4	514	1	AMY_BACAM
66	161.5	2.4	2413	1	PRO8_YEAST
67	161	2.4	1250	1	BXE_CLOBO
68	161	2.4	1608	1	HLVA_SERMA
69	160.5	2.4	1325	1	YDEK_ECOLI
70	160	2.4	1302	1	FRPA_NEIMB
71	159.5	2.4	1250	1	YFAL_ECOLI
72	159	2.4	704	1	CDGT_BACOH
73	159	2.4	984	1	HYSA_STRAG
74	159	2.4	1181	1	SCA2_STRPY
75	159	2.4	1196	1	BXCN_CLOBO
76	158.5	2.4	1295	1	BXA2_CLOBO
77	158	2.4	1122	1	ADPL_MVCGA
78	157	2.4	1286	1	ATDA_ECOLI
79	156.5	2.4	1290	1	BXCL_CLOBO
80	156	2.3	793	1	YHCD_ECOLI
81	156	2.3	1755	1	YJZ7_YEAST
82	155.5	2.3	537	1	ARP_PLAFA
83	155	2.3	905	1	HXA1_HAEIN
84	155	2.3	1167	1	SCA1_STRPY
85	154.5	2.3	1701	1	MSPL_PLAFM
86	154.5	2.3	2352	1	MOKC_SCHFO
87	154	2.3	1005	1	Y321_MYCPN
88	154	2.3	1035	1	NANA_STRPN
89	154	2.3	1196	1	AMYB_PAEPO
90	154	2.3	1755	1	XJZ9_YEAST
91	153.5	2.3	566	1	NPRE_BACCE
92	153.5	2.3	852	1	CSG_HALHA
93	153	2.3	967	1	Y5G0_CLOAB
94	153	2.3	1077	1	XINY_CLOMT
95	153	2.3	1228	1	SLAP_BACST
96	153	2.3	6669	1	NEBU_HUMAN
97	152	2.3	987	1	YD94_METJA
98	151.5	2.3	581	1	AMV1_SCHPO
99	151.5	2.3	703	1	CDGT_BACS2
100	151.5	2.3	718	1	CDGT_BACSS

ALIGNMENTS

RESULT 1

553 DTIKNSLHGLS-----DATN-----BWLDAIVHQSLADRENNSTENVVPIPNKSVFRAHD 603
 525 NKRLSLIATLPLEKADSNKEIISGLPEVITNLSNRSBGGKSERANMIFIRAH 584
 604 NNSQDQIONAIR-DVTGK-DYRFTTFDEQKSDAYIQDQNSYKKNYLNIPASVAILL 661
 585 SEVOTVIAKIAQINPKTDLGFTLDELKQAFKIFNEDMRQAKKYYTQSNPTAYALML 644
 662 TNKDTIPRYVGGYGLTGGQYMEHQRYVDYDTLNLKSKVKYVAGSQMOTWSVGNHNR- 720
 645 SNKDSITRLYGDMSYDQGYMAKTSQDYDAIDTLKARIKAAAGQDMKITYVEGDKSH 704
 721 -----ILTSRYGKGMAMPATGDTGDTGIGVGVVSNTPNLKLGVDKVVVLHMGAAH 773
 705 MDWDYTGVLTSRYGTCANEATDQGSSEATKTCGMVAVITSNPNLKLQNDKVINVMGAAH 764
 774 KNOQTRAAVLTGTVGVNINYSQDAP--VAMTBNGDLYLSSNVLVNVNKEERADYVQY 831
 765 KNOEYRPLLLTTKDGLTSYSDAAKSLYKTKNDKGLVFDASD-----IQGY 812
 832 ANPDVSYGLAVWPVGASDMDQARTAPSTPTEKNSGNSAYRTNAAFDNSVIFEPNFEVYTP 891
 813 LNPQVSYGLAVWPVGASDMDQVYVAASNAKANATQVYESSALDSQLIYEGESNFQDFV 872
 892 TKESERANVRIQANDPFASLGFTSFEMARQYNSKDRFLDSTIDNGAFTDRYDLGHS 951
 873 TKDSYTNKKIAQNVQLFKSGVTSFEMAPQYSSSEDSGLDSIIQNGYAFEDRYDLAMS 932
 952 EPKNYGTDEDLNIAIOALHKAGLQVMAWVDPDQIYNLPKGEVAVTVRVDGRGNVWKDAII 1011
 933 KKNYKSGQDMINAVKALHKSQIQVADNVDPQIYNLPKGEVAVTVRVDGRGNVWKDAII 992
 1012 NNNLYVNTIGGG-EYQKKYGGAFDLKLQKLYPEIFTKKQVSTGVADIPDSQKITEWSAKY 1070
 993 KNTLYAANTKSNQDGYQAKYCGAFSLSELAQKYPISFNRQTQISNGKKIDPSEKITANKAKY 1052
 1071 FNGTNILHRSYVYLKADGG-QYVNL-GTTTKOFLPQLNGEKKQNGEYVKGNDGN-VY 1127
 1053 FNGTNILRGVGVYVLKNDASDKYFELKNGT--YLPQMT--NKEASTGEV--NDNGHMT 1106
 1128 FYDLAGNVMKNTFIEDSVGNWYFDQDQKVMENKHFVDVDSYGEKGYTFPLKNGVSFRGG 1187
 1107 FYSTSGYQAKNSFYQDQKNGNYFYDNGHGWYGLQQLN-----GE--VQYFSLNGVQLRES 1160
 1188 LVOT-----DNGTYFEDNYGK-----MYRNOTINAGAMITYLDEN 1222
 1161 FLENADGSKNFTGHLGNRYSGNYSFNDNSKWRFYFDASGVMAVGLKTINGNTQYF--DQD 1218
 1223 GKLIKASY 1230
 1219 GYQVKGAW 1226

RESULT 2
 ID GTFC_STRM STANDARD; PRT; 1375 AA.
 AC P13470; P05427;
 DT 01-NOV-1988 (Rel. 09, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE Glucosyltransferase-SI precursor (EC 2.4.1.5) (GTF-SI)
 DE (Dextranase) (Sucrose 6-glucosyltransferase).
 GN GTFC.
 OS Streptococcus mutans.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 OC Streptococcus
 OX NCBI_TaxID-1309;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-GS-5;
 RX MEDLINE-89137980; PubMed-2976010;
 RA Ueda S., Shiroza T., Kuramitsu H.K.;

Sequence analysis of the gtfC gene from Streptococcus mutans GS-5.;
 Gene 69:101-109(1988).
 [2]
 RN SEQUENCE OF 1-349 FROM N.A.
 RP STRAIN-GS-5;
 RC MEDLINE-87308013; PubMed-3040685;
 RX Shiroza T., Ueda S., Kuramitsu H.K.;
 RA *Sequence analysis of the gtfB gene from Streptococcus mutans.*;
 J. Bacteriol. 169:4263-4270(1987).
 RL
 CC -!- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS. THAT ARE THOUGHT
 TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE
 OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIANE THE
 AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.
 CC -!- CATALYTIC ACTIVITY: Sucrose + {(1,6)-alpha-D-glucosyl}(N) - D-
 fructose + {(1,6)-alpha-D-glucosyl}(N+1).
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- DISEASE: DENTAL CARIES.
 CC -!- MISCELLANEOUS: GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA
 1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES
 WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH
 FORMS OF GLUCANS.
 CC -!- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-
 BINDING PROTEIN FROM S.MUTANS.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch)

 CC EMBL; M22054; AAA88592.1; ..
 DR EMBL; M17361; AAA88589.1; ..
 DR PIR; J0345; J0345.
 DR PIR; C33135; C33135.
 DR InterPro; IPR002479; CW_binding.
 DR Pfam; PF01473; CW_binding_1; 7.
 DR Pfam; PF02324; Glyco_hydro_70; 1.
 KW transferase; glycosyltransferase; signal; Repeat; Dental caries.
 FT SIGNAL 1 34
 FT CHAIN 35 1375
 FT DOMAIN 35 1050 GLUCOSYLTRANSFERASE-SI.
 FT DOMAIN 1126 1375 CATALYTIC (APPROXIMATE).
 FT DOMAIN 1126 1375 GLUCAN-BINDING (APPROXIMATE).
 FT DOMAIN 1126 1375 2.4 A. 1 C AND 1 AC REPEATS.
 FT REPEAT 1126 1159 A REPEAT.
 FT REPEAT 1169 1200 A REPEAT.
 FT REPEAT 1227 1238 C REPEAT.
 FT REPEAT 1253 1303 AC REPEAT.
 FT REPEAT 1318 1330 A REPEAT (INCOMPLETE).
 SQ SEQUENCE 1375 AA; 153022 MW; D4B80CBEE0AAE13 CRC64;

Query Match 35.38; Score 2347.5; DB 1; Length 1375;
 Best Local Similarity 41.3%; Pred. No. 3.2e-110;
 Matches 538; Conservative 181; Mismatches 408; Indels 175; Gaps 35;
 QY 19 SAVSGFDATIKLNDQYQALMGQLVLRFSKAADG--NPSGDNTVTDQFSKNVATGGN 76
 DB 29 TSUSGSLYKADSTDRQQAQYTESQASLVTTSEAKETLTDSTATSTATSTATVTDN 88
 QY 77 FDYVKNQNGVFEFSWHATNQSDDQSWIILVNGYKVKRQLVNDTKGGAAGFNNDVY 136
 DB 89 VS-----TTNQS-----TNTTANTA-----NFVV 107
 QY 137 KYNPAIENSSNGSGFGIITLPVTKNENNVQLVHRSNDVKTGEGNYV-----DFW 166
 DB 108 KPTTSEAKTNSDKIITTSKAVNR-----LPATGKFPANNNTAHPKTVI 154
 QY 187 SELMPVKDSFKNGPEKQ-----FGLQTINGQQYYIDPTTGPQRKNFL 231
 DB 155 DRIVTPKPKI-----GKLGKQPSLSQDDIAALGNVKNIRKNGYKYYY-KEDGTLOKNYAL 209

QY 232 QSCNNHLYEDSGVCTNALELOFAKGTYSSE---QYENGNAAYSYDDKSLIENYNGYLT 288
Db 210 NINGKTFEEDGALSNNTLPK--KGNITNDNNSAQAQYVSTOVAMFEHVDHYLT 267
QY 289 ADTWYRPKOLKDKTWTWDSKETDMRPIKLVWVWNTLTQAYLYLTKMQHGNLLPSALPFF 348
Db 268 AEWYRPKYLKDGTWQSTQKDFRPLMTWPDQETQRQVYVNNNAQLGHIHQ---Y 323
QY 349 NADADPAELNHYSEIYVQNYEKRIKSTGTGNTOMLRLYHDFVTVNPNWKNKSDSNVPSGIQ 408
Db 324 NTATSELOLNAQAOTQYIEEKITPAKNTNLRITISAFVTKQAWNSDSSEKPPDHLQ 383
QY 409 FOGGELKYENDLTPYANSYRLLGRMPN-----IKQTYRGQBELLANDIDNSN 459
Db 384 -KALYISNNKSLTQANSYRIINPTNQGKDPRTADRTTGGYEFLLANDVDNSN 442
QY 460 PVYQAEOLMWLYLLNFGTITANNQANFDSVRVDPADNIDADLMNIAODYNAYGM-D 518
Db 443 PVYQAEOLMWLHFMFNGNIYANDPANEFSIRVDADVNDVADLLQIAGDYKAAKGIHK 502
QY 519 SDVSNKHINILEDNHADPEYENKIGNPOLWMDPTIKNSLHGLSDATN-RWGLDAIVH 577
Db 503 NDAANDRLSILEAWSYNDTPYUDDGDMMNNDKRLSLIYSLAKPLNQSGMPLIT 562
QY 578 QSLADRENSTENVYIENYSFVRAHDNNSOQIQONAIR-----DVTGKDYHFTFEDEQK 632
Db 563 NSLVNRTDDNAETAAPVSYFIRAHDSVQOLIRNIRTEINPNVVG---YSFTTEEEK 619
QY 633 GIDAYIQONSTVKKYLYNIPASYAILLTKTIPRYVYGDLYTDGGYMBHORYDT 692
Db 620 AFEIYNKDLATEKYYTHYTALSYTALLTNKSSVPRVYIGDMFTDDGGYNAHKTINER 679
QY 693 LTNLLKSKVYVAGQSMOPVSGVGNMILTSVYKGGAWTADTDTGTRQGGIGVYVS 752
Db 680 IETLLKARIKYSVGOAMRQOV--GSEIITSVYKGGALKATDSDRTSGVAVIGS 738
QY 753 NTPMLKGVNKKVYLVHGAHKKQVRYANVLTDTGVINYTSQGP--VAMTDENGDIY 810
Db 739 NPSRLKASDVVNVNNGAAHKKQVRYANVLTDTGVINYTSQGP--VAMTDENGDIY 798
QY 811 LSSHLVYNGKEEADTAVQYANPQVSGVYLVWVPGVADNQDAPATPSTKNSGNSAYR 870
Db 799 FTA-----AD--IKGYPANPQVSGVYLVWVPGVADNQDAPATPSTKNSGNSAYR 846
QY 871 TNAAFDSNVIFAPSNVYPTTESERANVRLAQNAADFPASLGFTSFEMAPQYNSKORT 930
Db 847 QNAALDSRVMPGESNFQAFATKKEETNVVIAKNVYDKFAENGVTDFEMAPQVNSDGS 906
QY 931 FLDSITDNGVAFTRDYDLGMSPEPKYGTDEDLRNAQALHKAQLQVWADVPDQIYNLPG 990
Db 907 FLDSVIONGYAFTDRYDLGISKPNKYGTADLVKAIKALHSGKVKVWADVPDQIYNLPG 966
QY 991 KEVATVTRVDRGNWWDALINNLYVW--TIGGEVQKYGGAFLDKLQKLYPELFTK 1049
Db 967 KEVATVTRVDRGYTPVAGSQIKNTLYVVDGSKSGKQQAQYGGAFLEELQAKYPELFAK 1026
QY 1050 QVSTGVAIDPSOKITENSAKVFNGTINILHSGSGVYLKADGGQVYINLTGTTKQPLQLATG 1109
Db 1027 QISGVDPNPSVKIKQMSAKVFNGTINILHSGSGVYLKADGGQVYINLTGTTKQPLQLATG 1085
QY 1110 EKKQGNFVKG--NDG--NYFYDLAGNKNVNTFIEDSVG--NMYFFDQDQKVENKHFVD 1165
Db 1086 NPNHGTSSVTRGLVFDKGVYVYSTSGNAKNAPF--SLGNNWYFDNNGYMTVGAQSN 1143
QY 1166 VDSYGEKGYTFELKNGYFSGGLVQT-----DNGTY-----YFNGYK 1203
Db 1144 -----GANVIFLSNGIQLRNAIYDNGKVLSTYDNGDRYENGNYLFGQOQRYFN-GI 1196
QY 1204 MYRNQITINAGAMITFLDENGKLIKASNSDAEYPTSTVGKM 1245
Db 1197 MAVGLTRVHGAQVY-FDASG-----FOAKGQFITTAD-GKL 1230

RESULT 3
GTFB_STRMU STANDARD; PRT: 1476 AA.
ID AC P08987; 069381; 069384; 069387; 069390; 069396;
DT 01-NOV-1988 (Rel. 09, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Glucosyltransferase-I precursor (EC 2.4.1.5) (GTF-I) (Dextranucrase)
DE (Sucrose 6-glucosyltransferase).
GN GTFB.
OS Streptococcus mutans.
OS Bacteria; Firmicutes; Bacillus/Clostridium group: Streptococcaceae;
OC Streptococcus.
OC NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-GS-5;
RX MEDLINE-87308013; PubMed-3040685;
RA Shiroza T., Ueda S., Kuramitsu H.K.;
RT "Sequence analysis of the gtfB gene from Streptococcus mutans.";
RL J. Bacteriol. 169:4263-4270(1987).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-MT4239, MT4245, MT4451, MT4467, AND MT8148;
RX MEDLINE-98231643; PubMed-9570124;
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,
Kimura S., Hamada S.;
RT "Molecular analyses of glucosyltransferase genes among strains of
Streptococcus mutans.";
RL FEMS Microbiol Lett. 161:331-336(1998).
CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT
TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE
OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIANE THE
AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.
CC -1- CATALYTIC ACTIVITY: Sucrose + ((1,6)-alpha-D-glucosyl)(N) - D-
fructose + ((1,6)-alpha-D-glucosyl)(N+1).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- DISEASE: DENTAL CARIES.
CC -1- MISCELLANEOUS: GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA
1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES
WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH
FORMS OF GLUCANS.
CC -1- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-
BINDING PROTEIN FROM S. MUTANS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M17361; AAA8588.1; -
CC EMBL; D88651; BAA26101.1; -
CC EMBL; D88654; BAA26105.1; -
CC EMBL; D88657; BAA26109.1; -
CC EMBL; D88660; BAA26113.1; -
CC EMBL; D89977; BAA26119.1; -
CC PIR; B33135; B33135.
CC InterPro; IPR002479; CW_binding.
CC InterPro; IPR003318; Glyco_hydro_70.
CC Pfam; PF01473; CW_binding_1; 13.
CC Pfam; PF02324; Glyco_hydro_70; 1.
KW Transferase; Glucosyltransferase; Signal; Repeat; Dental caries.
FT SIGNAL 3 34 POTENTIAL
FT CHAIN 35 1476 GLUCOSYLTRANSFERASE-I.
FT DOMAIN 35 1051 CATALYTIC (APPROXIMATE).
FT DOMAIN 1097 1476 GLUCAN-BINDING (APPROXIMATE).
FT REPEAT 1097 1130 A REPEAT.
FT DOMAIN 1161 1470 5 X TANDEM REPEATS.
FT REPEAT 1161 1210 1.

FT REPEAT 1225 1275 2.
 FT REPEAT 1290 1340 3.
 FT REPEAT 1355 1405 4.
 FT REPEAT 1420 1470 5.
 FT VARIANT 62 65 S -> T (IN STRAIN MT4239).
 FT VARIANT 65 65 I -> T (IN STRAINS MT4239, MT4245, MT4251, MT4467 AND MT8148).
 FT VARIANT 68 68 A -> V (IN STRAIN MT4239).
 FT VARIANT 78 78 Q -> P (IN STRAIN MT4251).
 FT VARIANT 86 86 S -> I (IN STRAIN MT4239).
 FT VARIANT 89 89 S -> F (IN STRAIN MT4251).
 FT VARIANT 168 168 K -> N (IN STRAIN MT4251).
 FT VARIANT 276 276 D -> S (IN STRAINS MT4239, MT4245 AND MT4251).
 FT VARIANT 399 399 N -> R (IN STRAIN MT4239).
 FT VARIANT 474 474 I -> T (IN STRAIN MT4239).
 FT VARIANT 512 512 K -> R (IN STRAIN MT4239).
 FT VARIANT 519 519 F -> Y (IN STRAIN MT8148).
 FT VARIANT 701 701 T -> I (IN STRAIN MT8148).
 FT VARIANT 708 708 A -> V (IN STRAIN MT8148).
 FT VARIANT 938 938 F -> L (IN STRAIN MT8148).
 FT VARIANT 952 957 YGTPVA -> RGKPE (IN STRAINS MT4245, MT4251 AND MT8148).
 FT VARIANT 963 964 NT -> SV (IN STRAINS MT4245, MT4251 AND MT8148).
 FT VARIANT 968 970 VDG -> ADS (IN STRAINS MT4245, MT4251 AND MT8148).
 FT VARIANT 1086 1086 A -> T (IN STRAIN MT4239).
 FT VARIANT 1158 1158 S -> N (IN STRAIN MT4239).
 FT VARIANT 1163 1163 H -> Y (IN STRAIN MT4251).
 FT VARIANT 1168 1168 Y -> K (IN STRAIN MT4251).
 FT VARIANT 1182 1182 X -> C (IN STRAIN MT8148).
 FT VARIANT 1234 1234 A -> P (IN STRAIN MT4239).
 FT VARIANT 1263 1263 H -> R (IN STRAINS MT4239, MT4245 AND MT4251).
 FT VARIANT 1264 1264 H -> Y (IN STRAINS MT4245 AND MT4251).
 FT VARIANT 1272 1272 G -> S (IN STRAINS MT4245 AND MT4251).
 FT VARIANT 1329 1329 Y -> H (IN STRAINS MT4239, MT4245, MT4251 AND MT8148).
 FT VARIANT 1394 1394 H -> Y (IN STRAINS MT4245 AND MT4251).
 FT VARIANT 1402 1402 G -> S (IN STRAINS MT4245 AND MT4251).
 FT VARIANT 1459 1459 Y -> H (IN STRAIN MT4467).
 FT CONFLICT 570 570 R -> A (IN REF. 1).
 FT CONFLICT 800 817 AQQVRAASTPSTGK -> LKMFALRLRPHQQA (IN REF. 1).
 FT CONFLICT 1310 1310 H -> L (IN REF. 1).
 SQ SEQUENCE 1476 AA; 165685 MW; 3479B62B07694D98 CRC64;

Query Match 35.1%; Score 2330.5; DB 1; Length 1476;
 Best Local Similarity 42.8%; Pred. No. 2.5e-109;
 Matches 52; Conservative 185; Mismatches 379; Indels 135; Gaps 35;

QY 107 IVLVNGKVRQLVND-----TKGAAGFNNDVKYKPAIENSSK 147
 DB 35 LKADNSKESQISNDSTSVTANEESNVTETASKQEAASQTR--HTVTSSTSTSV 92
 QY 148 SGFGIITLPTVYKNVNLVHPSNDYKNGYVDFNSEL-----MPVKD 194
 DB 93 VNPKEVSNPYTV-GETASNGKLQQTTF-----VDKTEAANISKOTTEADTDV 146
 QY 195 SFQKNGPL--KQGLAVNGQVYIDPTGQPKNFLLQSGNNWYFD-----SDTGYG 247
 DB 147 DSNAANLQILEKLPVKEIDKYYVD--NGKVRTNFTLIAQDKILHFEDEGTAYTDS 205
 QY 248 TNALEAFAGKTSSNQYKNGNAYSYDDKSTENVNGYLTATWTRPKQILKDGTTWTD 307
 DB 206 TVNKDIV-----TTRSN-LYKYNQVDRSAQSFPHVDHYLTASSWTRPKYILKDGKT 260
 QY 308 SKETDMRPIILWVWPNLTQAYLVNWKQGNLLPSALPEENADAPAEALHYSELVQQN 367
 DB 261 STEKDFRLLTWMPDQETQYQVYVYMAQLGINKT-----YDTSNQLQNTIAAATQAK 316

QY 368 IEKRISGTGNTDLRLTLMHDFVTNNPWNKDSNVNPSGIQOGGFLKYEN--SDLTPEYAN 426
 DB 317 IEAKITTLKNTDMLRQTISAFVKTQSAHNSDSEK--PFDD--HLONGAVLYDNEGKLTPTAN 374
 QY 427 SDYELLGRMPIN-----IKDVTYRGQEFFLLANDIDNSFPVQAEQLNWLTYLLNFG 477
 DB 375 SNYKILNKPTNQTKGKDPRTADNTIGTQEFFLLANDVNSFPVQAEQLNWLHFLNFG 434
 QY 478 TIRANWDOANFDSVRVDAPDNIDADLMTAQDYFNAAYGM--D6DAVSNAKHINILSDNHA 536
 DB 435 NIYANDPANDFDSIRVDAYDNDYDADLLQIAGDYLRKAAGIKHKNDRKAANDBELSILEWSDN 494
 QY 537 DPEYFNKINPOLMDDTIKNLSNLHGLSDATN--RWGLDAIVHOSLADRENNSSTNVVPIPN 595
 DB 495 DTPVLUHDDGDMNIMONKRLSLLSLFLAQRSQWPLITNSLVNRTDNRRTAAPS 554
 QY 596 YSFVRAHNSQOIQNAIR-----DVTGKYHTFTFEDEQKIDAIYODQNSTVKKYNL 650
 DB 555 YSFIRAHDSVQDLIRDIKAEINPNVVG---YSETMEETKKAFFIYNKDLATKSKYTH 611
 QY 651 YNIPASVAILLTNKDITIPRVYVYGDLYTGGQYMEHOTRYVDITLNLKSRVKYVAGGQSM 710
 DB 612 YNTALSTALLTNTKSSVPRVYVYGDYDQYMAHKLINYEATLTKARIKIVSGGQMA 671
 QY 711 QTMVSQGNNTLSVRYGKGMATATDGTDTGTQIGVYVVSNTPLKLGVDNKVYLHMG 770
 DB 672 RNQOV--GNSIITSVRYGKALKATDGTDRITRTSGVAVIEGNPNPSRLKASDRVYVNMG 730
 QY 771 AAHKNQYRAVLTDTGIVNTYSDQAGP--VAMTDENGDIYLSHNLVYNGKEADTAV 828
 DB 731 AAHKNQYRPLLTDTGIRKAYHSDQEAAGLVRYTDRGELIFTA-----AD--I 778
 QY 829 QGYANPQVSGYLAIVVVPVAGSDNQDARTAPSTEKNSGNSAYRTNAAFDNSVIFAPSNFV 888
 DB 779 KGYANPQVSGYLVVVPVAGAAADQDVRAAATASTPDTGKSVHQAALDSRVMEGFSNFQ 838
 QY 889 YTPKESERANVRIAQADDFASLGFTSFEMAPQYNSKDRTEFLDSTDINGYAFTRDYDL 948
 DB 839 AFATKKEBYTNVYIAKNVDKPEAGVYDFEMAFYSSDSDGSLDSVIQNGYAFTRDYDL 898
 QY 949 GMSFPNKYGTDELRNATQALHAKAGLQWADWDVDPQIYNLPFGKEVATVTRVDDRGWVKD 1008
 DB 899 GISFPNKYGTADDLVKAIKALHSGIKVMADWDVDPQMYAFPEVVTATVDYKGTVPAG 958
 QY 1009 AITNNLVVN--TICGEYKYYKGAFLDKLQKLYPIETFKKQVSTGVVAIDPSQITENS 1067
 DB 959 SQIKNTLYVVDGKSSGKDQAKYGGAFLELOAKYLPFAFKQISTGVPMDPSVKIKOWS 1018
 QY 1068 AKYFNGTILHRSYGYVLKADG--QYVNLGTTTK--QFLPQLTGKKGQNGEVPKVGNDGN 1125
 DB 1019 AKYFNGTILHRSYGYVLKADG--QYVNLGTTTK--QFLPQLTGKKGQNGEVPKVGNDGN 1075
 QY 1126 YFFYDLACNMVKNFTIEDSVGNWYFFDQDGNVKNHNFVDVSYGKGTFFFLKNGVSEFR 1185
 DB 1076 YVYTSQYAKNTFISBG--DKWYIFDNGYVMTGAOSIN-----GVNTYFSLNGIQLR 1128
 QY 1186 GGLYCTDNGTV--YFQNGYKMYRN-----QTIWAGAM-----YVTLDEMGK 1224
 DB 1129 DAILKNEGTVAIYNGDGRYENGYFMSGVWRHFNNGEMSVGLTVIDGQVQYFDEMG-- 1187
 QY 1225 LKASYNDAEYPTSTDVGMK 1245
 DB 1188 -----YQAKGKFTVTD--GKI 1202

RESULT 4

ID STDF2_STRDO STANDARD; PRT: 1592 AA.
 AC P27470;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)

DE Glucosyltransferase-I precursor (EC 2.4.1.5) (GTF-I) (Dextranucrase)
 DE (sucrose 6-glucosyltransferase).
 OS Streptococcus downei (Streptococcus sobrinus).
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1317;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=6715;
 RA MEDLINE=9112327; PubMed=1704006;
 RA Abo H., Matsumura T., Kodama T., Ohta H., Fukui K., Kato K.,
 RA Kagawa H.;
 RA "Peptide sequences for sucrose splitting and glucan binding within
 RT Streptococcus sobrinus glucosyltransferase (water-insoluble glucan
 RT synthetase).";
 RL J. Bacteriol. 173:989-996 (1991).
 CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT
 CC TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE
 CC OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE
 CC AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.
 CC -1- CATALYTIC ACTIVITY: Sucrose + {(1,6)-alpha-D-glucosyl}(N) - D-
 CC fructose + {(1,6)-alpha-D-glucosyl}(N+1).
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- DISEASE: DENTAL CARIES.
 CC -1- MISCELLANEOUS: GTF-1 SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA
 CC 1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES
 CC WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-S1 SYNTHESIZES BOTH
 CC FORMS OF GLUCANS.
 CC -1- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-
 CC BINDING PROTEIN FROM S.MUTANS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: D90213; BAAL4241.1; -;
 DR PIR: A38175; A38175.
 DR HSP: P00695; 2HEE.
 DR InterPro: IPR002479; CW_binding.
 DR InterPro: IPR003318; Glyco_hydro_70.
 DR Pfam: PF01473; CW_binding_1; 16.
 DR Pfam: PF02324; Glyco_hydro_70; 1.
 KW Transferase; Glycosyltransferase; Signal; Repeat; Dental caries.
 FT SIGNAL 1 38 POTENTIAL.
 FT CHAIN 39 1592 GLUCOSYLTRANSFERASE-I.
 FT DOMAIN 39 1044 CATALYTIC (APPROXIMATE).
 FT DOMAIN 1093 1592 GLUCAN-BINDING (APPROXIMATE).
 FT REPEAT 1093 1142 6.5 X TANDEM REPEATS.
 FT REPEAT 1158 1207 1.
 FT REPEAT 1222 1272 2.
 FT REPEAT 1287 1337 3.
 FT REPEAT 1402 1451 4.
 FT REPEAT 1514 1563 5.
 FT REPEAT 1577 1592 6.
 FT REPEAT 1592 1592 7 (INCOMPLETE).
 SQ SEQUENCE 1592 AA; BC0A66D079451ECF CRC64;
 Query Match 34.2%; Score 2274; DB 1; Length 1592;
 Best Local Similarity 45.9%; Pred. No. 1.9e-106;
 Matches 496; Conservative 167; Mismatches 340; Indels 78; Gaps 29;
 QY 158 VTKRNNVQLVRRF SNDVKYEGEYVDFWSELMPVKDFSTQKNGPLKQFLQTLQYQY 217
 DB 116 VPTENEN---QGFTDEMLAEAKNVAESDSIP-----SDLAKMSNVKQVDCYY 163
 QY 218 IDFTQGRKKNFLQSGNNWYIFQSDTGV--GTNALQLQFAKCTVSSNEQ-YRNGNAAYS 274
 DB 164 YD-QDGNVKKNFASVSGDKIYIFD-ETGAYKDTSKVDADKSSSAVSQNTAFIAANNRAYS 221

QY 275 YDKSIEVNGYIADTWTRPAPKQILKDGTTWTDSKSTDMRPILMVWNPWLQAYLYNVM 334
 DB 222 TSKNFBAVDNITADSWYRPFKILKDGKTWTESGDDPRLPLMAWNPDTETKRYNVM 281
 QY 335 KQGNLPLSALPFFENADADPAELNHYSEIVQONIEKRISETGNTDMLRMLHDFVTNNFM 394
 DB 282 ----NKVVGIDKTYTAETSDQDLTAARAEVQARIEKITSNTKWLRAISAFVKTQPO 337
 QY 395 WNKDSNVNFGSQIOGQGLUKYEN-SDLTPYANDVRLGLRMPINIKDO-----TYR--- 445
 DB 338 WNGSEKPYDD--HLQNGALLFDQDLTPTQSNYRLLNRTPTNQTGLDSRFTYPNPD 395
 QY 446 --GQEFLLANDIDNSNPVQAEOLNWLXYLLNFGYITANNDOANEDSVYDAPNIDAD 502
 DB 396 PLGGYFLLANDVNSNPVQAEOLNWLXYLLNFGYITANNDOANEDSVYDAPNIDAD 455
 QY 503 LMTIAOYFNAAYGMD-SDAVSNKHNILEDNHADPEYFNKIGNPQLTMDTITKNSLH 561
 DB 456 QLQISSDYLAAYGIDKNNKNNHNSIVENASDNDTPYLHDPGDNLNMMNKKFRLSMLW 515
 QY 562 GLSDATN-RNGLDALVHQSLADRENNSTENVIPNYSFVRAHNNNSQDOIQNAIR-DVWG 619
 DB 516 SLAKPTDVRSLNPLHNSLVDRVDDREYETVPSFARAHSEVQDIIRDIKAEINP 575
 QY 620 KDY-HTFTFDEQKIDAYIQDQNSTVKYNLYNIPASVAILTNKDTIPRVYGLDLYD 678
 DB 576 NSFQSYTQBEIDQAFIYNEDLKKSDKKYTHVNPVLSYLLLTNKGSPRVYVYGDMDTD 635
 QY 679 GGQYMEHQTRYDPLTNLLSKRVYVAGGSMQTMVGGNNNLTISVRYCKGKAMTATDYG 738
 DB 636 DGGYANKNTYDAIESLLKARKRYVAGGSMQTMVGGNNNLTISVRYCKGKAMTATDYG 694
 QY 739 TDETRTQIGVGVVSWTFNLKLVNDKVV-LHMGAAHKNQYRAAVLTFTDGTGVINYSDOG 797
 DB 695 DATRISGCVGVVQNGQNFSL--DGKVVALLMGAHANQYRALVSTKGVATYATDAD 752
 QY 798 AP----VAMTDENGDLXLSHNLVVGKEADTAVQGYANPDVSGYLAVVWVPGASDND 853
 DB 753 ASKAGLVKRTDENGYLFLNDL-----KGVANPQVSGELQVWVPGVGAADDOD 800
 QY 854 ARTAPSTKNSGNSAYRTNRAFTDSNVLFPAFSPNFTPTKESRANVRITAOADPFASLG 913
 DB 801 IRVAASDTASDGLKSLHQDAMD SRVMEFGFSNQSPATKEEYTVVIANVVDKFSWG 860
 QY 914 FTSEMAPQYNSKSDRTPLDSTIDNGYAFTRDYLGMSEPNKYGTDEDLRNAIQALHKG 973
 DB 861 ITDEMAPQYVSVSTGGQFQDSVIQGYAFTRDYLGMSEPNKYGTDEDLRNAIQALHKG 920
 QY 974 LQVWADWYDQIYNLPGEVATVTRVDGRGVKDALINNNLYVNT-IGGGEYQKYYG 1032
 DB 921 LKVMADWYDQIYNLPGEVATVTRVDGRGVKDALINNNLYVNT-IGGGEYQKYYG 980
 QY 1033 AFLDKLQKLYPEITKKQVSTGVAIDFSQKITEMSAKYFNKTNILHRGSGYVLKAD-GGQ 1091
 DB 981 AFLDKLQKLYPEITKKQVSTGVAIDFSQKITEMSAKYFNKTNILHRGSGYVLKAD-GGQ 1040
 QY 1092 YVNLGTTTKQYPLQTLGKQKQNGEYVKGNGEYIPYD--LAGNWNKTFIEDSVGNWY 1149
 DB 1041 YFNASDT-LFLPSLLGKVVESGIRY----DKRGYIYNSATGDOVKASAFITEA-GNLY 1094
 QY 1150 FFDQDQKVKENKHFVDVDSYGEKGTFFLKNVSGFRGLVQTD--NGTYFDNKGVRN 1207
 DB 1095 YFGKGYVNVQAQTIN-----GANYFFLENGTALR-NTIYDQAGNSHYANDGKRYN 1147
 QY 1208 Q 1208
 DB 1148 E 1148

RESULT 5

GTFI_STRDO

ID GTFI_STRDO

STANDARD;

PRT; 1597 AA.

AC P11001;
 DT 01-JUL-1989 (Rel. 11, Last Created)
 DT 01-JUL-1989 (Rel. 11, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE Glucosyltransferase-I precursor (EC 2.4.1.5) (Gtf-I) (Dextranucrase)
 DE (Sucrose 6-glucosyltransferase).
 CN GTFI.
 OS Streptococcus downei (Streptococcus sobrinus).
 CC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 CC Streptococcus;
 CC NCBI_taxid=1317;
 OX [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN-MFE28;
 RC MEDLINE=87308014; PubMed=3040686;
 RA Ferretti J.J., Gilpin M.L., Russell R.R.B.;
 RA "Nucleotide sequence of a glucosyltransferase gene from Streptococcus
 RT sobrinus MFE28";
 RL J. Bacteriol. 169:4271-4278(1987).
 CC -!- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT
 CC TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE
 CC OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIANE THE
 CC AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.
 CC -!- CATALYTIC ACTIVITY: Sucrose + [(1.6)-alpha-D-glucosyl](N) - D-
 CC fructose + [(1.6)-alpha-D-glucosyl](N+1).
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- DISEASE: DENTAL CARIES.
 CC -!- MISCELLANEOUS: GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA
 CC 1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES
 CC WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH
 CC FORMS OF GLUCANS.
 CC -!- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-
 CC BINDING PROTEIN FROM S.MUTANS.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 DR EMBL; M17391; AAC63063.1; -
 DR InterPro; IPR002479; CW-binding.
 DR InterPro; IPR003318; Glyco_hydro_70.
 DR Pfam; PF01473; CW-binding_1; 19
 DR Pfam; PF02324; Glyco_hydro_70; 1.
 KW transferase; Glycosyltransferase; Signal; Repeat; Dental caries.
 FT SIGNAL 1 38
 FT CHAIN 39 1597
 FT DOMAIN 39 1050
 FT DOMAIN 1099 1597
 FT DOMAIN 1099 1597
 FT REPEAT 1099 1132
 FT REPEAT 1163 1213
 FT REPEAT 1227 1277
 FT REPEAT 1292 1342
 FT REPEAT 1352 1399
 FT REPEAT 1406 1455
 FT REPEAT 1465 1512
 FT REPEAT 1519 1568
 FT REPEAT 1582 1597
 SQ SEQUENCE 1597 AA; 177080 MW; B9P86A200868798E CRC64;

Query Match 34.2%; Score 2271; DB 1; Length 1597;
 Best Local Similarity 46.2%; Pred. No. 2.7e-106;
 Matches 500; Conservative 155; Mismatches 344; Indels 84; Gaps 29;
 QY 158 VTVKNENVOLVHRRF---SNDVNTGSGNVDFWSELSMPVKDSFQKGLQGLQINGQ 214
 DB 122 VPTENQVFTDEMLAEAKVATAESNIP--SDL-----AKMSNVKQVDGK 166

QY 215 QYIIDPTGQPRKFLQSGNNWIYFDSDTGV--GTNALELQFAKGTWVSSE--QYRNGNA 271
 DB 167 YYYD-QDGNKKNFAVSGEKIYYFD-ETGAYKDTSKVREADKSGSDISKEETFAANR 224
 QY 272 AYSDYDKSIENGVYLPADTWYRPQILKQDTTWDKETDMRPFILAVWPNVLTQAYYL 331
 DB 225 AYSTAENFEADINLTADSWYRPKSLKQKWTESKDDFELLAMWPDTEIRKRYV 284
 QY 332 NYMKQHGMLLPALPFFNADADPAELNHYSEIVQNTKRISEGTNGTDLRLTLMHDPVN 391
 DB 285 NYM----NKVVGIDETTYTAETSQADLTAAAEVQARTEOKITTEONTKWLREALSAFVK 340
 QY 392 NPMWNKSENVMNFSIOFGGFLKYEN-SDLTPYANSYRLLGRMPINIKDQ-----TYR 445
 DB 341 QPQWNGSEKPYDD--HLQNGALKFDQSDLPDTPDQSNRYRLNKTPTNOTGSLDSRSTYN 398
 QY 446 -----GQEFLLANDINSNPVQAEQLNMLYLLNCTLTANNQANFDSVRVADPNI 499
 DB 399 ANDPLGGYELLANDVNSNFVQAEQLNMLYLLNCTLTANNQANFDSVRVADPNI 458
 QY 500 DADLNIADQYFNAAAGMD-SDAVSNKHINLEDNHADPEYFNKIGNPOLMTDDTIKNS 558
 DB 459 DADLQISSDYKAAAYGDKNNKANNHVSIVEAWSMDTPYLHDDGDNLMWMDKPRLS 518
 QY 559 LNHGLSDATN-RWGLDAIVHQSADRENNSTENVVPIWYSPVRAHDNNSQDQIQATR-D 616
 DB 519 MLWSLAKPLDKRSGNLPLHNSLVVDREVDRETVETPSFARAHDFVODLTHDIIKAE 578
 QY 617 VTCKDY-HTFTFEDQKGDIAITQDNSTVTKRYNLYNPASYAILLTKNKTIPRVYIGDL 675
 DB 579 INPNAFGYSFTQDEIDQAFKYNEDLKKTKKTYHNVPLSYTLILLNKGSIPIVYIGDM 638
 QY 676 YIDGQYMEHOTRYTDTLNLKSKVYVAGGQSNQVMSVCGNNILTSVRYGKATAT 735
 DB 639 FUDGQYMAKNTVNDATLTKKARKYVAGGQAMONYQI--GNGEILTSVRYGKALKQS 697
 QY 736 DTGDTETRTGIGVVVSVNTPNKLGVNDKVV-LMGRAHKNQYRAVLTDTQDVVNTS 794
 DB 698 DKGDATRTSGVVMGQPNFSL--DGKVVALLNAGAAHQAQYRALMYSIKDGVAYAT 755
 QY 795 DGAP-----VAMTDNGDLYLSHNLVNGKEADTAVOGYANPDVSGYLAWVPVGASO 850
 DB 756 DADAKAGLVKFTDENGILYFLNDGL-----KGVANPQVSGFLQWVPVPGAAD 803
 QY 851 NQDARTAPSTEKNSGNSAVRTNAAFDNSVIFAFSNFYVTPTEKSERANVRIADNFEA 910
 DB 804 DQDIRVAASDTASTDGKSLHQDAANDSRVMEGFSNFQSPATKEEYTNVVIANNVDFV 863
 QY 911 SLGFTSFENAPYNSKDRFTLDSTIDNGYAFTRDYLDGNSEPNKYGTDEDLRNAIQALH 970
 DB 864 SWGITDFEMAPQVSVSTDQGLDLSVIQNGYAFTRDYLDGMSKANKYGTADQLVKAIALH 923
 QY 971 KAGLOVADWVPOULNLPGEKAVTAVTDORGNNVKDAIINNLYVNT--IGSGEYQKK 1029
 DB 924 AKGLKVMADWVPOMTFPKQEVVTVTRDKFGKFIAGSQINHSILYVTDTKSGDDYQAK 983
 QY 1030 YGGAFLDKLQKLYPELFTKKQVSTGVAIDPSQKTEPWSAKTFNGTNIILHRSGYVIAKDG 1089
 DB 984 YGGAFLDELKELYPELFTKKQISTGQADPSVKIKOWESAKYFNGSNILRCADYVLSQA 1043
 QY 1090 GQYVNGTTKKQFLPIQLGCKKQGNESFVGKNDQNYFYD--LAGNMVKNTIEDSVGN 1147
 DB 1044 SNKLVNSDCKLFLKTLKLGQVVEGIRF----DGTGVVNSSTTGEKVTDSFITEA--GN 1098
 QY 1148 WYFFDDQKGVNKHVFDVDSYGEKGT-YFFLKNQVSPRGGIVQD--NGVYVYDNNGKM 1204
 DB 1099 LYTFQDQGVAVTGAQNI-----KGSNYFLANGAALR-NVYTTDAQQNHQYNGDGR 1150
 QY 1205 VRN 1207
 DB 1151 YEN 1153

Qy	233	--SGNNWIIYDSDTGVGNALBL-----QFAKGTVSSNEQYANGNAIYSD-----	276
Db	588	ATSKNYWIGVSAIFDQCEGAMSAPKPIPVNWGAQAPSAGKYNNGNATGYFOLSWKAVS	647
Qy	277	-----DKSTENWNGYLTADTWYRPAQILKDGTTTDSKRETDMPILVMWPNTL---TQA	328
Db	648	GATGYKVQVFNKGSETLD-----LGNQTSWT-----TWGKXI-----WPTSAEIKAGK	691
Qy	329	YLLNYMKQHGNLLP--SALPFPNADADPAELNHYSEIVQON-----	367
Db	692	YALHLKDGSGAELPIMPPTYKNAGDGAKRNYSPFKIAYNKDGEALASPAATPALPDIA	751
Qy	368	-----IEKRISETG--NTDWLRT-----LHMDFFVNN	392
Db	752	RPKNWGYLYTNKSSQGTGYNLWLEKVNQNAKGVKNYIYNGKEYQSPFDVGDAHWTQNK	811
Qy	393	PMNKDSENVPFGIOF-----QGGFLKYEISDLTPYANSDYR-----L	431
Db	812	NIW-PTSEERKAGYKLTDDGGGSELALDPSPVYNNANGNYKGGKNYSFTLVAYDANGET	870
Qy	432	LGRMPINIKDYYRGQEFLLAND-----IDNSNPVQABQLANWLYLLNFGTITANNDQAN	487
Db	871	IPTAPFN--PTPHEGAFFLGTBEYSIIID-----IPSSQLANGA-----TCNVIYNEDLS	918
Qy	488	FDS-----VRVDAPDNI-----DADLMNIAQDYFNAAAYGMSDAYSNNKHIN--	528
Db	919	IDGGPGLGLSRTYNSLSDSHLFGQGYADAETSVIDT--GGANYIDEDATTHRETKK	976
Qy	529	-----ILEDNHADP-----EYFNKIGNPOLTMDDDTIKNSLNIGLSDAT	567
Db	977	ADGTYOPPTGVLELTETADQFLTKTDQTNAYFNKGGKL-----	1017
Qy	568	NRGLDALVHQSILADRENNSTENVIPNVSFVRADHNSDQIQNAIRDVTGKDYHFTFY	627
Db	1018	-----QKVDDGHNAQT-----VYTYNKKQL--TATYDASGRKL--TFTY	1053
Qy	628	ED-----EQKGIDAY-----IQDNSTVKKYNYLN--IPASVAILLTWKT	666
Db	1054	DENGHVTSITGPKNKKVTVSYRNDLLKKVTDGCTVTSDYDVSDEGLRVQYVSAMSTEAKP	1113
Qy	667	IPRVYGDLYTDGQYMEHQR-----YYD-----TLTNLKSRYKVIVAG	706
Db	1114	VFTFY-----QYSGHRLKELAINAKETVYSYDADKKTLMTQNGRKYQYGYNEA	1164
Qy	707	GQSMQTSVSGVGNNNILTSYRGKGAMTATGTDGTRTQGCYGVVVSNTPNLKGVDNKKV	766
Db	1165	GNPIQVIDAEGKLITNWKY-----EGNNVEDYDPN--DVGT-----	1201
Qy	767	LHWGAHKWQYRAAVLTTTDDGVHNTSDQAPVAMTBENGDLLSHNLVVGKEADT	826
Db	1202	---GKATESYQY-----KDKG--NVTSVKDAYGTEYB-----YKNNDVTKMKDTSBNV	1246
Qy	827	AVOCYANPDVSGYLAVWPVPGASDQARTAPSTEKNSGNSAYRT--NAAFDNSWI---F	881
Db	1247	TDIAYDGLDA-----VSETDQSGKSSAAVYDXGNQIQSKDLASNTNLKDGSF	1297
Qy	882	EA-PSNFPVYPTKESRANVRAQIADNAPFASLGFTSFEMAPQYNSKDRFTLDSITDNGY	940
Db	1298	EAKSGSNLTAASDRKRKISV-IADKSGVLS--GSKALEVLQSQTS-----AGTDHGY	1346
Qy	941	AFPTDRYDLGMSBN-----KYGTD-----EDLRN-----AIGALHK--AGLOVM	977
Db	1347	SSAQOTVEL--BENTYTLGSKIKTDLAKSAYFNIDLRDKDKRIOWIHNEYSALAGK	1403
Qy	978	ADWYPDQI--YNLP---GKEYATV---TRVDDRGWVKDAIINNNLVYVNTIGGCETQKK	1029
Db	1404	MDWTKRQITFTTANAKGAWYMEVDHDKDKGKRAWFDEV-----QLEKGEYSSS	1454
Qy	1030	Y-----GGAFDLKLQKLYPEI-----FTKKQV-----	1051
Db	1455	INPVQNSFTSATENNVNVSASVDSEEGFNDVSLKAAATSNASQAGSVTKOTVILGOSAN	1514

```

1052  QY  -----STGVADPSOKITE-----WSAKYFNGTNILHRGSGYVL 1085
      ||:|:| | | | |
1515  Db  KDPVYLTLTGSKASSYKFFDEKDYSLQANVYADGTGIYNAKFPSTQPNRAAYVIP 1574
      :|:|:| | | | |
1086  QY  KADGGQYVNLGTTTKQPLQLTGKQKQNEGFVKGMGDNVYFYDL----AGNMVKNTFI 1141
      || || | | | | | | | | | | | | | | | | | |
1575  Db  K-----TK-----PI-----NKVDISILFQKSATGVTWEDDIRLEGSLLTKSTY- 1614
      || || | | | | | | | | | | | | | | | | | |
1142  QY  EDSVGNWYF-----FPDQKMKVENKHFVDVDSYGEKGTYYFF-----L 1178
      || ||:|:|:| | | | | | | | | | | | | | | | |
1615  Db  -DSNGNYVTKFEDEBLGYATSTDYDETKKTS-----ETDAKGEKTTTYDQADQLTNMIL 1668
      || || | | | | | | | | | | | | | | | | | |
1179  QY  KNGVSFPGGLVQTDNGYVYFDNNGVKWNRNOTINAGA---MYITLDENGKLIKASYNSD-- 1233
      || || | | | | | | | | | | | | | | | | | |
1669  Db  SNGTSI-----LHSDKKEGNEV-SKTIIRAGADQYKFEIDVHGKLVK---TTDPL 1714
      || || | | | | | | | | | | | | | | | | | |
1234  QY  -----ABYPTSTDVGRMLDON 1249
      :|:|:|:| | | | | |
1715  Db  GNVLASEYDANSNLTKTISP 1735
      || || | | | | | | | | | | | | | | | | | |

RESULT 8
TOXB_CLODI STANDARD; PRF: 2366 AA.
ID AC P18177;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Toxin B.
GN TOXB OR TCDB.
OS Clostridium difficile.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1496;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-VPI 10463;
RX MEDLINE-90336540; PubMed-2374729;
RT Barroso L.A., Wang S.Z., Phelps C.J., Johnson J.L., Wilkins T.D.;
RT "Nucleotide sequence of Clostridium difficile toxin B gene.";
RN Nucleic Acids Res. 18:4004-4004(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-VPI 10463;
RX MEDLINE-92293124; PubMed-1603068;
RT Eichel-Streiber C., Laufenberg-Feldmann R., Sartingen S.,
RA Schulze J., Sauerborn M.;
RT "Comparative sequence analysis of the Clostridium difficile toxins A
and B.";
RN Mol. Gen. Genet. 233:260-268(1992).
CC CC -1- DISEASE: CLOSTRIDIUM DIFFICILE PRODUCES TWO TOXINS, AN
CC CC ENTEROTOXIN CALLED A AND CYTOTOXIN B.
CC CC -----
CC CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC CC EMBL; X53138; CAA37298.1; -
CC CC EMBL; X92982; CAA63562.1; -
CC CC EMBL; X60984; CAA43299.1; -
CC CC PIR; S10317; S10317.
CC CC InterPro; IPR002479; CW_binding.
CC CC Pfam; PF01473; CW_binding_1; 18.
CC CC Cytotoxin; Toxin.

```

```

1052  QY  -----STGVADPSOKITE-----WSAKYFNGTNILHRGSGYVL 1085
      ||:|:| | | | |
1515  Db  KDPVYLTLTGSKASSYKFFDEKDYSLQANVYADGTGIYNAKFPSTQPNRAAYVIP 1574
      :|:|:| | | | |
1086  QY  KADGGQYVNLGTTTKQPLQLTGKQKQNEGFVKGMGDNVYFYDL----AGNMVKNTFI 1141
      || || | | | | | | | | | | | | | | | | | |
1575  Db  K-----TK-----PI-----NKVDISILFQKSATGVTWEDDIRLEGSLLTKSTY- 1614
      || || | | | | | | | | | | | | | | | | | |
1142  QY  EDSVGNWYF-----FPDQKMKVENKHFVDVDSYGEKGTYYFF-----L 1178
      || ||:|:|:| | | | | | | | | | | | | | | | |
1615  Db  -DSNGNYVTKFEDEBLGYATSTDYDETKKTS-----ETDAKGEKTTTYDQADQLTNMIL 1668
      || || | | | | | | | | | | | | | | | | | |
1179  QY  KNGVSFPGGLVQTDNGYVYFDNNGVKWNRNOTINAGA---MYITLDENGKLIKASYNSD-- 1233
      || || | | | | | | | | | | | | | | | | | |
1669  Db  SNGTSI-----LHSDKKEGNEV-SKTIIRAGADQYKFEIDVHGKLVK---TTDPL 1714
      || || | | | | | | | | | | | | | | | | | |
1234  QY  -----ABYPTSTDVGRMLDON 1249
      :|:|:|:| | | | | |
1715  Db  GNVLASEYDANSNLTKTISP 1735
      || || | | | | | | | | | | | | | | | | | |

RESULT 8
TOXB_CLODI STANDARD; PRF: 2366 AA.
ID AC P18177;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Toxin B.
GN TOXB OR TCDB.
OS Clostridium difficile.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1496;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-VPI 10463;
RX MEDLINE-90336540; PubMed-2374729;
RT Barroso L.A., Wang S.Z., Phelps C.J., Johnson J.L., Wilkins T.D.;
RT "Nucleotide sequence of Clostridium difficile toxin B gene.";
RN Nucleic Acids Res. 18:4004-4004(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-VPI 10463;
RX MEDLINE-92293124; PubMed-1603068;
RT Eichel-Streiber C., Laufenberg-Feldmann R., Sartingen S.,
RA Schulze J., Sauerborn M.;
RT "Comparative sequence analysis of the Clostridium difficile toxins A
and B.";
RN Mol. Gen. Genet. 233:260-268(1992).
CC CC -1- DISEASE: CLOSTRIDIUM DIFFICILE PRODUCES TWO TOXINS, AN
CC CC ENTEROTOXIN CALLED A AND CYTOTOXIN B.
CC CC -----
CC CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC CC EMBL; X53138; CAA37298.1; -
CC CC EMBL; X92982; CAA63562.1; -
CC CC EMBL; X60984; CAA43299.1; -
CC CC PIR; S10317; S10317.
CC CC InterPro; IPR002479; CW_binding.
CC CC Pfam; PF01473; CW_binding_1; 18.
CC CC Cytotoxin; Toxin.

```

SQ SEQUENCE 2366 AA: 269709 MW; E1024BD8A56ADF CRC64;

Query Match 3.34; Score 222; DB 1; Length 2366;
Best Local Similarity 20.4%; Pred. No. 0.0022;
Matches 220; Conservative 155; Mismatches 401; Indels 304; Gaps 60;

QY 270 NAAYSYDKSIENVNGYLTAQI-----WYRPKQILKGGTTWTDSK-----ETDMRRIL 317
DB 1459 NIPYSFVD--SEKENGFTNGSTKEGLFVSELPDVVVILSKVYMDSDSKPSFGYYSNNLKDVK 1517
QY 318 MYWNPEN--TILQAYILNMYKQHGNSLPSALPFPFNADADPAELM--HYSEIVQONIEKRISE 374
DB 1518 VITKDNVILGYLK-----DDIKISLSLTQDEKIKLSVHLDSGVAEILKFNRR 1571
QY 375 TGTNDWLTLLHDFVTPNPMNKNDSNVNFSIQPGQGLKYENSDLSPYANSYVRLAGR 434
DB 1572 KGNWTSDSLM--SFL-----ESMNIKSI-----FVNFLQSNIKFIILDAFIIISGT 1615
QY 435 MPINIKQDTYRGQEFLLANDIDNSNPVQAQLNLWXYLLNFGTITAND-----484
DB 1616 TSI-----GQFEEDENDNIQ-----YFIKFTLETNTYTLVGNRQMI 1656
QY 485 -QANFDSVRVDAPNDIDALMIAIDYFAAYGMDSDAVSNKHI---NILEDMNHADPEY 540
DB 1657 VEPNYD---LDSGDISSTVINFESQYL---YGIDS--CVNKKVISPNIYDEINITPVY 1708
QY 541 FNKIGNPQLTMDDT--IKNSLNHGLSDATNR--WGLDAIVHQSILADRNNSNTENVIPNY 596
DB 1709 ETNNTYPEVILVDANYINEKINVININDLSIRVNSNDGNDFILMSTSEENKVSQVKI---1765
QY 597 SPVRAHDNNSQDQIONARIDVTGKHHTFTPEDEKGDIDAYIQDQNSVKKVNLNIPAS 656
DB 1766 -----RNVVFKDKTLANKLSFNFSDKQ-----DVPVS 1793
QY 657 YAILLTKDITPRVIYGD-----LYTD-----GGQMEHOTRY--DT 692
DB 1794 EIL-----SFTPSYEDGLIGDLYSLYNEKPYINFGMVSGLIYINDSLYFPP 1848
QY 693 LYNLKSRY-----KY-----VAGGOSMTMSVGGNNILTSVRYGKGAWTADTGTDFTR 743
DB 1849 VNNLTGFTVGDYKYPENINGAA---SIG--ETIIDDKNYVFNQSGVLQTVGFSTE 1902
QY 744 TOGIGVVSNT--PNLK--LGVDKVVLEHGAHKNQXRAAV--LTTTGDVINTS--DQ 796
DB 1903 DGFKFAFANTLDENLEGEADFTGKLIDENIYTFDNYRGAVFWKELDGMHIFSPET 1962
QY 797 GAPVAMTDENG--LYLSSHLVNGKEADTAVQYANPDVS--GYLAVVVPVPGASDQ 852
DB 1963 GRAFKGLAQIGDYKYFNSDGVYMQKGFVSINDNKHIFDDSGVMKVGYTEI-----DGK 2015
QY 853 DARTAPSTEKNSCNSAYRTNAAFDSNVIFAFSPNVTPTKESERANVRIQADFPASL 912
DB 2016 HFYFAENGEMQIG-----VFNTEDGFKPAHINEDLGN-----BEGSEISYS 2057
QY 913 GFTSEMAPQVNSSKDRNFLDSTIDNGVAFDTRDYDL-----GMSEPNKYGTDEDLRNAIQ 967
DB 2058 GILNFN-----NKIYFDDSFYAVVGWKLDEGSKYIFDEDTAEAY- 2098
QY 968 ALHKAGLOVMADWVPDQIYNLPG--KEVATVTRVDDRGVNMKDA-----IINNLLYVV 1018
DB 2099 -----IGLSLIND--GOQYFNDGIMQVGFVT--INDKVYFSDSGIIESGVQNIIDNNYFI 2151
QY 1019 NWIG--GGEYQKKYGGAFDLKLQKLYPEIFTPKKQVSTGVAIDPS--OKITBWSAKYPNG 1073
DB 2152 DDMGIVQIGVFTSDGYKYFAPANTVNDIY-----GOAVEYSGLVRVGE--DYIYFGE 2203
QY 1074 TNLHRSQGYVLKADGGVYINLGTTKQLPLQLTGKKGQNGEFGVKDNGNYXYFDLAG 1133
DB 2204 YTIETGNLYDNESDKYI-----FNP-----ETKACKGILNIDDIKIFYF--DEKQ 2249
QY 1134 NMVKNTFIEDSVGNVFFDQDKM-----VENKHFDVDSYGEKGTFFFLKNGVSFRGG 1187

DB 2250 --IMRTGLISFNNYFNENGEMQFGYINTEDKMFY-----FGEDGV-----MQIG 2294

QY 1188 LVQDNGTYFYDNYGKMYRN---QTINAGAMIY---LDENKGLIKASYNDAEYPTST 1240
DB 2295 VFNPDPGPKYFAHQNTLNDENPEGESIN-----YTGWLDLDE-----KRYFTD-EYIAAT 2343

RESULT 9

ID APU_THETU STANDARD; PRT; 1861 AA.

AC P38536;

DT 01-OCT-1994 (Rel. 30. Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 01-MAR-2002 (Rel. 41, Last annotation update)

DE Amylopullulanase precursor (Alpha-amylase/pullulanase) (pullulanase type II) [Includes: Alpha-amylase (EC 3.2.1.1) (1,4-alpha-D-glucan glucanohydrolase); pullulanase (EC 3.2.1.41) (1,4-alpha-D-glucan glucanohydrolase) (Alpha-dextrin endo-1,6-alpha-glucosidase)].

GN AMTB.

OS Thermoanaerobacter thermosulfurogenes (Clostridium thermosulfurogenes).

OC Bacteria; Firmicutes; Bacillus/Clostridium group;

OC Thermoanaerobacter group; Thermoanaerobacterium.

OX NCBI_TaxID=33950;

RP SEQUENCE FROM N.A. EML;

RC STRAIN=DSM 3896 / EML;

XX MEDLINE=94252998; PubMed=8195085;

RA Matuschek M., Burchhardt G., Sahn K., Bahl H.;

RT "Pullulanase of Thermoanaerobacterium thermosulfurogenes EML (Clostridium thermosulfurogenes): molecular analysis of the gene, composite structure of the enzyme, and a common model for its attachment to the cell surface.";

RT J. Bacteriol. 176:3295-3302(1994).

RL -I- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-alpha-glucosidic linkages in oligosaccharides and polysaccharides

CC -I- CATALYTIC ACTIVITY: Hydrolysis of (1->6)-alpha-D-glucosidic linkages in pullulan and in amylopectin and glycogen, and the alpha- and beta-limit dextrins of amylopectin and glycogen.

CC -I- SUBCELLULAR LOCATION: CELL-BOUND. IT C-TERMINUS MAY SERVE AS AN S-LAYER ANCHOR.

CC -I- PTM: GLYCOSYLATED.

CC -I- SIMILARITY: BELONGS TO FAMILY 13 OF GLYCOSYL HYDROLASES, ALSO KNOWN AS THE ALPHA-AMYLASE FAMILY.

CC -I- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.

CC -I- SIMILARITY: CONTAINS 3 S-LAYER HOMOLGY (SLH) DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; M57692; AAB00841.1; -

DR HSP; Q08751; IBVZ.

DR InterPro; IPR000461; Alpha_amylase.

DR InterPro; IPR003961; FN_III.

DR InterPro; IPR001119; SLH.

DR InterPro; IPR004185; alpha-amylase_N.

DR InterPro; IPR004193; Isoamylase_N.

DR Pfam; PF00128; alpha-amylase; 1.

DR Pfam; PF02806; alpha-amylase_C; 1.

DR Pfam; PF02903; alpha-amylase_N; 1.

DR Pfam; PF00041; fn3; 2.

DR Pfam; PF02922; isoamylase_N; 1.

DR Pfam; PF00395; SLH; 3.

DR SMART; SM00060; FN3; 1.

DR PROSITE; PS01072; SLH_DOMAIN; 3.

DR Hydrolase; Glycosidase; Carbohydrate metabolism; Signal; Repeat;

KW Multifunctional enzyme; Glycoprotein.

FT SIGNAL 1 35

```

731 AMTATDGTGDTETGGVGVVVSNPNLK---LGW-----ND 763
    :| |: | |: |||| |::
1465 GNSSSSTSGSSTGSTITSNTISNTISNTITNGKVTTFLDGAKADLVNKKDK 1524
QY   764 KVLHMGAAHNKOYRAAVLTDTTGVTNYSDQAPVANTDENGDLVLS----SHNLVVN 819
    ||| |: | |: :::: |:: |: | |: | |: | |: | |: | |: | |: | |: |
1525 KVVFDDITTIGEGQQ---KVQISKXLDILTSAANKCKIVIKSDNASIALTYDALNQIQIN 1581
QY   820 GKEDATAVOCYAMPDYGSYLAVNVPV-----GASDMQDARTAPSTEKNSGNSAYRTNA 874
Db   1582 G- - - VMSIKNKGKPNTVNYSLNVDIIISGISGNVTLAKPVEVTLN-----ISKA 1631
QY   875 FDSNVIFPAFNFVYTPTKES-----BRANVRIQAONDAFFASLAGTFSEMAPQYNSS 926
Db   1632 NDRPKV- - - AVYYNPNTNQWEYVGKVDASSGTITFNATHFSQ--YAAFEYDKFTNDI 1685
QY   927 KD-----RIFLSTDINGYAFTRDRLGMKEPNKYGF-----D 959
    || |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |
Db   1686 KONMAKDVIELASHRHIVEGMIDHQY-----EPNKTVTRAFTAMILRLNIKD 1734
QY   960 E-----DLR-----NAIQALHKAGL 974
    ||: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |
Db   1735 ETYSGESDVKSGDWYANAIEAAYKAGI 1762

RESULT 10
APU_THESA
AC APU THESA STANDARD; PRG: 1279 AA.
IC P36905;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Amylopullulanase precursor (Alpha-amylase/pullulanase) {Includes: Alpha-amylase [EC 3.2.1.1] (1,4-alpha-D-glucan glucanohydrolase); Pullulanase [EC 3.2.1.41] (1,4-alpha-D-glucan glucanohydrolase)}
DE DE (Alpha-dextrin endo-1,6-alpha-glicosidase)}.
GN APU.
OS Thermoanaerobacter saccharolyticum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OX Thermoanaerobacter group; Thermoanaerobacterium.
OK NCBI_TaxID=28895;
XX [1]
NN SEQUENCE FROM N.A.
RP STRAIN=BGA-R1;
RX MEDLINE=94161525; PubMed=8117096;
RA Ramesh M.V., Pedkovyrov S.M.; Lowe S.E., Zeikus J.G.;
RT "Cloning and sequencing of the thermoanaerobacterium saccharolyticum BGA-R1 apu gene and purification and characterization of the amylopullulanase from Escherichia coli.";
RN Appl. Environ. Microbiol. 60:94-101(1994).
RL [2]
RR IDENTIFICATION OF PROBABLE VECTOR CONTAMINATION.
RP Robison K.;
RA Unpublished observations (NOV-1994).
RL
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-alpha-glucosidic linkages in oligosaccharides and polysaccharides.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of (1->6)-alpha-D-glucosidic linkages in pullulan and in amylopectin and glycogen, and the alpha-and beta-limit dextrins of amylopectin and glycogen.
CC -1- SIMILARITY: BELONGS TO FAMILY 13 DF GLYCOSYL HYDROLASES, ALSO KNOWN AS THE ALPHA-AMYLAZE FAMILY.
CC -1- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN BY THE PRESENCE DF AN EXTRA C-TERMINAL SEGMENT DF 9 RESIDUES THAT SEEMS TO ORIGINATE FROM A PUC-TYPE VECTOR.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/or send an email to license@isb-sib.ch).
```

CC EMBL; L07762; AAA19800.1; ALT_SEQ.
DR HSP; Q08751; 1BZ.
DR InterPro; IPR000461; Alpha_acylase.
DR InterPro; IPR003961; FN.III.
DR InterPro; IPR004185; alpha-acylase_N.
DR Pfam; PF00128; alpha-acylase; 1.
DR Pfam; PF02806; alpha-acylase_C; 1.
DR Pfam; PF02903; alpha-acylase_N; 1.
DR Pfam; PF0041; fn3; 2.
DR SMART; SM00060; FN3; 2.
KW Hydrolyase; Glycosidase; Carbohydrate metabolism; Signal; Repeat;
KW Multifunctional enzyme.
FT SIGNAL 1 35
FT CHAIN 36 1279
FT DOMAIN 929 1017
FT DOMAIN 1156 1248
FT ACT_SITE 629 629
FT ACT_SITE 658 658
FT ACT_SITE 735 735
SQ SEQUENCE 1279 AA; 142430 MW; 095CCBCA391624DD CRC64;

Query Match 3.38; Score 216.5; DB 1; Length 1279;
Best Local Similarity 20.28; Pred. No. 0.0017;
Matches 269; Conservative 166; Mismatches 491; Indels 407; Gaps 74;

QY 31 TNDQYQALNGOLVLLRESKAADGNPSGDNFTVQPSKNYATTTGNGPDYKVGNGVQVEFS 90
DB 79 TPTQLKAGSYQYKVALNHSWGGVPSGGLNTLNTDSYVFW--FDY---NTQSVTDS 133
QY 91 GWHATQNSDKDSOWIYVLVN---GKE-----VKRQLVNDTKEGAAGFN---RNDV-Y 136
DB 134 TKY-TPISENKLPRLVGTIQAIGAGKMDPCTSTAIMDNDNVDYTAHIFKGYQY 192
QY 137 KV--NPAIENSSGCGQITLPTVYKNEWQVLRPSNVDKGT---EGNYVDFWSEL 190
DB 193 KVTGLTWAENYGANGVQ-----DGSNIQL---SVANDADITFFYDANTHNTWNTYS 241
QY 191 PVKDSFGKNGPLQFGLQTINGQYY-----IDPTGQP-----RKN 228
DB 242 PT-----LTGLDNIYYDLDLKHDTDFEFPNPFCAIKVGQVTLRIQAKN 286
QY 229 FLQSGNNHIYFSDTGVGTNALIEFAKGVTSSENGYRNGNAAYSDDKSIENWNGVLT 288
DB 287 HDLESA-RISYWD--INKRTLEPWRIGESPDGNYEWEIKLSFDHPT----- 333
QY 289 ADTAYRPQILKDGTTWTDKSDRPLIYAVWNTVTOAYLLMYKQHGNNLPSALPFF 348
DB 334 -RIWY--YFILLKGT-----KTAYYGNDDQLGQ-LGKATDTV 367
QY 349 NADADPAELNHYSEIVQONIEKRISGTGNDLRLMHDFTVNNPMNKDSERNVFSGIQ 408
DB 368 MKD--*FELTVYDKNEDTN-----DWMKGYWYQIFDPDRFNGDTSNDHAYTLS 413
QY 409 FQGGFLKYSNDLTPYANSYRLLGRMPINKDQTYRGQBFLLANDIDNSNPVYVQABOLN 468
DB 414 RGNDPIEFHN-----DWNLDLPDNPNNAGTPGYT-GDIWSNDFGGDLKGDIDKLD 463
QY 469 WL-----YLLNFGSITANDQANPDSVRVDPADMDADLMIADQYNAAYGMDSDA 521
DB 464 YLKGGLGVSYVLYNFIPEPSNKHKYDTADYTKID-----EMFGTTQDFEK-----LMSDA 512
QY 522 VSNKHINILED--WNHA--DPEYFNKIGN-PQLTMDDTIKNSLNHGL-----SDA 566
DB 513 HA-RGINILLDGVNHNFSDDSIYPRNRVCKYFDLCAYQDWKDG-NQSLSPYGDWNTINS 570
QY 567 TNR--WGLDAI-VHQSLADRENSTE--NVVIPNYSFVRAH-----DNNSQD-----QI 610
DB 571 TYECWGVYSLPVLKSLNGSEYNTSNANFIINDKNAISKYWLNPDENLNDGAGWELDV 630
QY 611 QNAIR-----DVTGKDYHTF----- 625

DB 631 ENEVAHDFTWTHFRDAINTVKPEAPWIAENWGDASDLLDGSFNSVANYQFRNDIIDFLIG 690
QY 626 --TFED---EOKGIDAYIQDQNSTVKYKLYNIPASYAI--ILLNRDITIPRYVYGLDLYDG 679
DB 691 QSFDDGNGQHNPIDAAKLQDR-LMSIYERYPLPAPYSTMNLGSHDTM-----SILTVF 743
QY 680 QOYMHQTRYDTLTNLLKSRVYVAGQSQMOTMSVCGNNILTSVRYGKAMTATDTGT 739
DB 744 GYNSADPNENSDDAAKRLAEQKLKLA---TILOMGYFG---MADIYVGBAGVSGCKDP 795
QY 740 DETRT-----QGIGYVSVNTPLKIGVNDKVVVLLHMGRAHKNOQYFAA--VL 793
DB 796 DDRFTFWGNETALQDFKVVSSIRNNNQVLATG--DLEFLYA-----QNDVYAIGRRII 849
QY 784 TTTDGVINTSDQAPVAMTDENGDLXLS-----SHNLVWNGKEADTAVOGYA 832
DB 850 NGKDAFGNSYFDSAIAIVAINRNSDQIITIDTTKFLRDGVAFKDLING-DKSYTINGGQI 908
QY 833 N---PDVSGYLAHVVPYVGCASDNQDARTAPSTEKN-----SGN-----SAYRTNAAFDSNVI 880
DB 909 TINIPAMSGVWMLI-----SDDGQDL-TAPQVPSNVVATSCNGKVDLSWSQSDGATGYNIY 962
QY 881 FEAFSNFYITPFKESERANVRIQONADFFASLGTSTFEMAPQYNSSKDRTFLDSTIDNGY 940
DB 963 RSVYEGGIYE-----KIASNV-----TGTTFEDTNTVNTGL 992
QY 941 AFTDRIYDLGSEPNKYTDEDLRNAIQALHAKGLQVNMADWYVDQIYNLPQKEVATVTRVD 1000
DB 993 ----KYVYAIASVDELGNESSEM--SIDTVAYPAYPI--GWVGNLTQVVVDNHVLSVSNPTE 1044
QY 1001 D-RGNVWKDAIINNLYVWNTIGGSEYKKGGAFLDKLQ-KLYPEITFKQVSTGVAID 1058
DB 1045 DIYAEVWADGLTNTSGGPNMIAQLGY-KYVGTVNDSVYGSVYNSVY-----GVD 1094
QY 1059 PSOKITEWSAKYFN--GTN-----ILHKGSGYVLKADGGQYVNLGTT----- 1098
DB 1095 DSD-FTWVNAQYVGDIGNNQYKASLHLNRSMGYLAHFSNDQGSWTTTIDTSLFYVPS 1153
QY 1099 ---TKQFLPI-----QLTGEKKQNGEFGVKGNDGNYFYDLAG-----N 1134
DB 1154 DDLIKTAPILNQPVESSRVSITWSPSTDNVGIY-----NYEIRSDGDTFNKIAVSN 1208
QY 1135 MVKNTFTEDSVGN 1147
DB 1209 EYVN-YVDTSVIN 1220

RESULT 11
PMFC_PROMI ID PMFC_PROMI STANDARD; PRT; 828 AA.
AC P53514;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DB Outer membrane usher protein pmfC precursor.
GN PMFC.
OS Proteus mirabilis.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Proteus.
OX NCBI_TaxID:584;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-HI4320;
RX MEDLINE-95047519; PubMed-7959033;
RA Massad G., Mobley H.L.T.;
RT "Genetic organization and complete sequence of the Proteus mirabilis
RL pmfC fimbrial operon."
RC Gene 150:101-104(1994).
CC -!- FUNCTION: INVOLVED IN THE EXPORT AND ASSEMBLY OF PMF FIMBRIAL
SUBUNITS ACROSS THE OUTER MEMBRANE.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane
(By similarity).
CC -!- SIMILARITY: BELONGS TO THE FIMBRIAL EXPORT USHER FAMILY.

EMBL; X62680; -, NOT_ANNOTATED_CDS.
DR Ecogene; BglI307; ydBA.
KW Hypothetical protein; Complete proteome.
FT CONFLICT 489 I -> V (IN REF. 2).
FT CONFLICT 495 495 I -> V (IN REF. 2).
SQ SEQUENCE 2003 AA; 205949 MW: B83A12C8B53220EE CRC64;

Query Match 3.2%; Score 210; DB 1; Length 2003;
Best Local Similarity 19.4%; Pred. No. 0.0069;
Matches 276; Conservative 166; Mismatches 508; Indels 470; Gaps 70;

QY	55	NPSGDNVTVD-----QFSKNVAT-----RG-----GNFDYKVNMGQVERSG	91
DQ	497	NEGESTITNGTGTQINGNDATANNSGKTIVDGKDSTGKIAGNIGVLNLSLTVTG	556
QY	92	WHATNQSD-----KDSOMILVIYVNGKEVKRQLVNDTKREGAAGFRNRDVIYKV	139
DQ	557	ABGVENIGDNTVNNKGDIVVSDTGISIGVLNG-----EGATVSNTGDVNVSN	604
QY	140	PA-----IENSMSGFQGIITL-----PVTVKNEYQLVHRFSDNYK-RGE	179
DQ	605	EATGSIITNSGKYSLAGSMVGDPSTGVDLNGNNSVTLAAKLDKLVGQKATGINVSD	664
QY	180	GNYDFWSELAPVXDSFKGNGPLKQGLQTINGQQYIDPTG---OPRKNFLOSGNN	236
DQ	665	ANTYNITGNVLVDK-----TADNAEYFFDSPSVGINVYGSDNNVTLDGKL	711
QY	237	WIYFSDRT-----GVG-----FNALQLQFAKGVSSNEQYRNCG	270
DQ	712	TVSDSEVTSRQSNIJFGSABKTSGLVIGDGNVTVMNGGLELIGEKNALDGSQVTSLR	771
QY	271	AAYSVDKSIENVG-----YLTDATVPYRPQOTIKDGTWTDSKETDMRPILMVMPNTLT	326
DQ	772	TGYSY--TSVIVVSESSVYLNGDT-----TIS	797
QY	327	QAYILNY-----MKQHNLPSALPFNFADAPAEUNHSEIVQOIEKRISETGNTDMLR	382
DQ	798	GEPLFGAGVIRVQDKALL-----ETSGSATLTMDQIDSFHHGTRTVEIQ	843
QY	383	TLMHDFVTNNPMNKDSNVNFGIO-FGG-----ELKYENSOLTFPVANSDYRL	432
DQ	844	NLGFAFVI-----GENTYGIINGSTISLLQNGKDPAPSIFVLATNGSGATNAGTIKGYT	898
QY	433	GRMPIKIDQTYRGQEFLANDIDNSNPVQAQLNWLYLLNFCTITANNDOANFDSVR	492
DQ	899	EQHSVFNKYSTGNSFIENDV-----SSITGLVAQSNTII-936	
QY	493	VDAFPNDIDADLMTAQDYFNAYGM-----DSAVNSKHINILEDNHADPEYFNKIGNPQ	548
DQ	937	-----NTDSGHIDL---YGRGSVGMALATAEAENQGXITLDSMRYDA-----NDT	979
QY	549	LTMDDTIKNS-LNHGL-----SDATNRGLDAIVHQSLADRENNSTEVVV	592
DQ	980	TAMRDIASNSAIDGTGVGVTDSYGAGKNATAINQLGGVITIYNAGAGMAAYGASNTV	1039
QY	593	IPNYSFVRAHDNNSQDQIQNAIRDVTGRDYHTFTFEDEQKGI-----DAYIQDQNST	644
DQ	1040	I-NQGTIMLEKNGYDDSLAANTLVGMAVYEHGSTAINDOTGVININVGTGQAFYNDGTGT	1098
QY	645	VKKY-----NLYNTPASY-AILLNKKOTIPR-----	669
DQ	1099	IVNYGTICTFGVCOSGSEYNNTTDFSILYTGGDTITRSGETVTILNKSAADVTKLAGNVV	1158
QY	670	---VYGDLYTDGQYMESHQTRYDYTLTNLKSRVYKYGQSGMQTSMVGGNNILITSVR	726
DQ	1159	NSGLTSGDIITVSSGLENTSG-GIINNVLKDKGAVIKNAGVMTNNVDVSGGITLINA-	1215
QY	727	YKGAMTATDT---GTDRTYQGIGVVYSNTPNLKLGYNKAVVLHMGAHKNOQYRAAVL	783
DQ	1216	--GEMTAQITMNAGADSS-----LVNNT-----GTINKIVQNAGVFNSSGVTGRMM	1260
QY	784	TTDGVINYTSQGPVAMTMDENGLYSSHLNVVNGKEADTAVQGVANPDVSGYLAWV	843


```
Db 1170 LQPCGIESRVTLNWSPSADDDVAIFGYEIKYSSSTGPFKIAFVS--DSVYVVDV 1227
Qy 1012 NNILVYVMTIG-GEYOKKYKAGFLDKLQKLYPIPFYKKQVSTGVAI---DPSOKITEWS 1067
Db 1228 NGVYVYKVAVDTSYNKTASNTV-----KATPDIPIK-VTFNVNTPDYTPD----- 1274
Qy 1068 AKYFNGTNIL-----HRSGGVVLKADGGQYVNLGTTTKQFLPIQLTGKKGQNEGFV 1119
Db 1275 ----DSVNIAGNFPDFAWPNANOMTKA-GSNTVSYLTNLNEGTOI-----EYKARGSWD 1325
Qy 1120 KGNNGNYFYDLAGNMVKN-----TFIEDSVGNV-----YFFDQDGKRWENKH 1162
Db 1326 KVEKGEY-----GNEIDNRKITVNGSNTWNVNDTVQWRWDVPIYIYSPKDKTIVDAN 1379
Qy 1163 FVDVDSYCEKTYEFLK---NGVSFRGLVTDNGTY 1196
Db 1380 TSEIEIKG--NTYRGAKVTYNDESF---VQOENGVF 1410

RESULT 16
P2P_LACP
ID P2P_LACP STANDARD; PRT; 1902 AA.
AC Q02470;
DT 01-JUL-1993 (Rel. 26, Created)
DE 01-JUL-1993 (Rel. 26, Last sequence update)
DE 15-JUL-1998 (Rel. 36, Last annotation update)
DE PII-type proteinase precursor (EC 3.4.21.96) (Lactocepin) (Cell wall-
DE associated serine proteinase) (LFI51).
GN PRTP.
OS Lactobacillus paracasei.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
OC Lactobacillus.
OX NCBI_TaxID:1597;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN-NCDO 151;
RC MEDLINE-92381481; Pubmed=1512565;
RA Holck A., Naes H.;
RT Cloning, sequencing and expression of the gene encoding the cell-
RT envelope-associated proteinase from Lactobacillus paracasei subsp.
RT paracasei NCDO 151.*;
RL J. Gen. Microbiol. 138:1353-1364(1992).
RN [2]
SEQUENCE OF 189-196.
RX MEDLINE-92226694; Pubmed=1564442;
RA Naes H., Nissen-Meyer J.;
RT Purification and N-terminal amino acid sequence determination of the
RT cell-wall-bound proteinase from Lactobacillus paracasei subsp.
RT paracasei.*;
RL J. Gen. Microbiol. 138:313-318(1992).
CC -1- FUNCTION: PROTEASE WHICH BREAKS DOWN MILK PROTEINS DURING THE
CC -1- GROWTH OF THE BACTERIA ON MILK.
CC -1- CATALYTIC ACTIVITY: Endopeptidase activity with very broad
CC specificity, although some substrate preference have been noted,
CC e.g. large hydrophobic residues in the P1 and P4 positions, and
CC Pro in the P2 position. Best known for its action on caseins,
CC although it has been shown to hydrolyse hemoglobin and oxidized
CC insulin B-chain.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Cell wall.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
CC SUBTILASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M83946; AAA25248.1; -.
DR PIR; B44858; B44858.
```

```
DR HSSP; P00782; 1S01.
DR MEROPS; S08 019; -.
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR003137; PA.
DR InterPro; IPR002029; Peptidase_S8.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR Pfam; PF02225; PA; 1.
DR Pfam; PF00082; Peptidase_S8; 3.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00136; SUBTILASE_ASP; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; FALSE_NEG.
KW Hydrolase; Serine protease; Cell wall; Zymogen; Signal;
KW Transmembrane.
FT SIGNAL 1 33 POTENTIAL.
FT PROPEP 34 187 POTENTIAL.
FT CHAIN 188 1902 PII-TYPE PROTEINASE.
FT DOMAIN 188 1876 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1877 1895 MEMBRANE ANCHOR (BY SIMILARITY).
FT DOMAIN 1896 1902 CYTOPLASMIC (POTENTIAL).
FT ACT_SITE 217 217 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 281 281 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 620 620 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DOMAIN 1867 1872 CONSERVED IN GRAM-POSITIVE COCCI SURFACE
FT PROTEINS.
SQ SEQUENCE 1902 AA; 200253 MW; D8C9F38CEE5DA582 CRC64;

Query Match 3.0%; Score 201; DB 1; Length 1902;
Best Local Similarity 18.8%; Pred. No. 0.018;
Matches 276; Conservative 184; Mismatches 531; Indels 474; Gaps 73;

Qy 12 VYPSLYNSAVSGFTTIKLTNDQYQALNGQLVLLRFSKAADGNPSGDNVTYD----- 64
Db 316 VFTNSDTSATTGSATLSAIEDSAKIGADVLNMSL-----GSDSGNOTLEDPEIAAVQ 368
Qy 65 -----QFSKNVATTGGNFYVKVNGNQVEFSGWHATN-----QSNQD 100
Db 369 NANESGPAAVISAGNSGTSGSATQGVNRYVGLQDN--EMVPTGTSRGATTVASAENTD 426
Qy 101 KDSQWIIIVLYNGKEVKRQLVNDT-----KEGAAGFNNDVYKYNPAIENSS-----MS 148
Db 427 VISQ-AVTITDGKDL--QLGPETIQLSSNDFTGSFDQKKFYVVKDASGDI.SKGAADYTA 483
Qy 149 GFQGIILPVTYVKNVQVLRFSNDVYKTEGNYVDFWSELMVPVKDSFGKNG----- 201
Db 484 DAKGKIAI-----VKRGELAFAD-----KQYQAAGAAGLIIV 517
Qy 202 -----PLKQ-----FGLQTINGQQYIIDPTTGQPRKNFLLQSGNNWYFSDTGC 245
Db 518 NNDGTATPLSIRLTITFTFTFGLSSKTGQKLV-----DWTAPIDDS 559
Qy 246 VGTIN-ALELOFAKGTVSSNEQYRNGNAAYSIDDKSIENVNGYLTADTYRPFKQLKDGTT 304
Db 560 LGVKIAL-----TLLPNQKYT-----EDKMSDFTSYGVPVSNLSFKPDITAPGGNI 604
Qy 305 W-----TDSKETDRPILWVWFWPLTQA-----YLYNMQHGMELPSAFPEN 349
Db 605 WSTQNNNGYTNMSGTSMAFPFIAGSQALLKQALNNKNPNFYADYKQLKGTALADFLKTV 664
Qy 350 AD-ADPAELAHYISIVQONTSEKRISETGNTDMLKTLMHDFVTNPNMKNKSENVNESGIQ 408
Db 665 MNTAQPIINDINYNVI---VSPRQAGGLVDVKAAT--DALEKNP-----STVV 708
Qy 409 FQGGFLKYENSDLTPYANSOYRLILGRMPINIKQTYNGQEFLLANDIDNSNPVVAQRLN 468
Db 709 AENGYPAVELKDFTS-TDKTFKL-----FTNRTTHELYQMSNDTINAVYTSATDPN 761
Qy 469 --WLYYLLNFGTITANDQANFDSVRVDAFNDIADLMTAQDYFNAY----- 515
Db 762 SGVLY-----DKKIDGAAGKAGSDITVPAGKTAQIEFTLSLPKSFQQQOFVE 808
```

QY 516 -----GMSDAVSNKHINILEDNHAD-PEYFNKI-----GN-----POLYMDDTKNS 558
DB 809 GFLNFKSGDRLNLYMGFGFGDNDGKIVDSLNGITSPAGNIGVPLTLTKNTKNTQHY 868
QY 559 LNHGLSDATNAGLDALVHOSLA-DRENNSTENVVIPNYSFVRAHDNNSODQIQ---NAI 614
DB 869 YCGWTDADGQTYD---DOAIFSSDKNALYNDJSMOYILLRNISNVOYDILDGQGNKV 925
QY 615 RDVTKGYHTTFEDEQKIDAYIQDQNSVTVKYKLNINPASYAILLNTKNDTIPRYVYG 674
DB 926 TTESSSTNQTKTYDAHS-----QKTYINAPAWDQTYDQD-----GN 965
QY 675 LXT-DGQYMEHQRYFYDTLNLKSKVKY-AGQSQNTQSVSGN-NNILTSVRYKGA 731
DB 966 IKTADGSI-----TYRISGVEGDKKQVDFPFKLDKSKAPTVRH----- 1006
QY 732 MTATGTGDTETRTQ-----GIGVVVSNTPNLKLGVNDKVLHMGAAHKNQOYR 779
DB 1007 -VALSAKTENGKTQYLTABAKDDLGLDA---TKSVKTAINE-VTNILDATFTDAG-- 1057
QY 780 AAVLTITDGVINYR-----SDQAPVAMTDENGDIYL-----SS 813
DB 1058 ----TTADG---YTKIEPLSDEQAQALNGDNSAELYLTDMASNATNQDASVQKPGSTS 1110
QY 814 HNLVYNG---KEEADTAQGY-ANPDVSG-YLAVWVPVPGASDQDARTAPSTKNSGNSA 868
DB 1111 FDLVINGGGIPDKISSTTGTTEANTQSGTYFSGTYPAAVDG--TYTDAQGKHDLNWT 1168
QY 869 YRTNAFDSNVIFRAPSNEVTPKE--SERANVRIQAADFFASLFTSEMAPQYNSS 926
DB 1169 Y--DAATNSFTASMAVTNADYAAQVDLYDAKHTQLLKHFTKVLRTAPTDL-KFNNG 1225
QY 927 KDRFLDSTI-----DNGYAFTRDYDLGMEPNKYG-----TDEDL 962
DB 1226 SDQT-SEATIKVGTQVSSDTKTVNVGDTVAALDAQHPSVDPVYVNDTITKVTATDEG 1284
QY 963 RNAIQALHKAQVMAWVPPQIYNL-----PGKEVATVT-- 997
DB 1285 NTTTTQ-----KTITSSYDPLKNAVNTFDQGVKGFANEFNATSAKFDYDKTGATITGK 1339
QY 998 -----RVDRGNVWDAIINNLYVNTIGGEY-QKYGGAFLDKLQ-KLYPE--I 1045
DB 1340 VKHPTTLTQVDS-----KQISIKNDLTFSEFIDLTLGQKPFQGVVGGDTTQNTQFQAL 1394
QY 1046 FTKKQVSTGVALDPSQKITEWSAKYFNTNHLRSGVYKADGGQYNNLGTETKQFLPI 1105
DB 1395 FILDVAFTSLDSDTD---APVYNDPNF-----QITGTATDRAQILSL 1436
QY 1106 QLTGKKGQNGEFGVAGNDQNYFYDLAGNMVKNFTIEDSVGNWYFFDQDGKKVKNKHEVD 1165
DB 1437 AINGSHVASQYADINISGK-----PGHMA-----IDQPVKLLGSKNVL 1476
QY 1166 V---DSYGEKGTFFLKNVSGFRGLVQTDNGTYTFDNYGKVRNQTNAGANIYTLDEN 1222
DB 1477 VAVTDS--ENNT-----TTKKITVYE-----PKTLAAPTVPSTPEP 1513
QY 1223 GLKIKASYNDAEYPT---STDVGK 1244
DB 1514 AKTYTLTANAATGETVQYSDGK 1538

RESULT 17
ID OMPB_RICRI
AC Q53047: STANDARD; PRT: 1654 AA.
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Outer membrane protein B precursor (168 kDa surface-layer protein)
DE (Surface protein antigen) (Cell surface antigen 5) (Scas) (rOmpB)
DE (rOmp B) [contains: 120 kDa surface-exposed protein (surface protein
GN OMPB. (120 kDa outer membrane protein ompB); 32 kDa beta peptide].

OS Rickettsia rickettsii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsia; Rickettsia.
OX NCBI_TaxID=783;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-R;
RX MEDLINE=92167802; PubMed=1724278;
RA Gilmore R.D. Jr., Cieplak W. Jr., Policastro P.F., Hackstadt T.;
RT "The 120 kilodalton outer membrane protein (rOmp B) of Rickettsia
RT rickettsii is encoded by an unusually long open reading frame:
RT evidence for protein processing from a large precursor.";
RL Mol. Microbiol. 5:2361-2370(1991).
RN [2]
RP SEQUENCE OF 279-1654 FROM N.A.
RC STRAIN-R;
RX MEDLINE=90136087; PubMed=2515418;
RA Gilmore R.D. Jr., Joste N., McDonald G.A.;
RT "Cloning, expression and sequence analysis of the gene encoding the
RT 120 kD surface-exposed protein of Rickettsia rickettsii.";
RL Mol. Microbiol. 3:1579-1586(1989).
CC -1- FUNCTION: THE 120 kDa SURFACE-EXPOSED PROTEIN IS A MAJOR
CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.
CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-
CC LAYER WITH HEXAGONAL SYMMETRY.
CC -1- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMP FAMILY.
CC
CC THIS SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; X16353; CAA34403.1; .
CC InterPro; IPR003858; rOmpA_rOmpB.
DR Pfam; PF02708; rOmpA_rOmpB; 1.
KW Antigen; S-layer; Cell wall.
FT CHAIN 1 1333 120 KDA SURFACE-EXPOSED PROTEIN.
FT CHAIN 1334 1654 32 KDA BETA PEPTIDE.
FT DOMAIN 1181 1188 POLY-THR.
FT SEQUENCE 1654 AA; 168184 MW; D7AB70FB7087F618 CRC64;
QY 15 SLVNSAVSGFDTTTLKNDQYQALNGQLVLLR--FSKAADGNPSGDNVTYDQFSKNYAT 72
DB 337 SVDNGKATIDGGVYAKDMVYQSANATGQVNFRIYDVGDGT-TAFKTAASKVTITQDS 395
QY 73 TGGNFDY-----VKVNGNQVFESGHWATNQSKDKDSQWIIVLVNGKEVRQLVNDTKEG 136
DB 396 MFGNTDFGNLAAQIKV-PNALTGTGNETGDSNPGTAGVITFD-----ANGLES 445
QY 127 AAGFNRDVYKVPAINENS-----SMSCFQGIITLPVTVKXENVOLVHRFSND---VKTGE 179
DB 446 ASA-----DANVAVTNNTATETASG-AGVVQLSGTHAAE-----LRLGNAGSIFKLAD 492
QY 180 GNYVDFWSELMPVKDSFKGNGPLKQ-----TGLQTINGQOYYIDP----- 220
DB 493 GTVI-----NGKVNQALVGGALAAATITLTDGSAITIDGIGNAGGAAL 536
QY 221 ---TTGQPRKNFLIQSGNNWYIFDSDTGVGTNALELQFAKGTVSSNQYRN---GNAAY 273
DB 537 QRITLANDAKKTLTLGKANII-----GAGGCIIDLQANGGTIKLSTQNNIVYVDFDLAI 590
QY 274 SYDDKSIENYNGYLTADTYWRPKQILKDGTTWTSKEDMRPILMVMWPNLTQAYVLYN 333
DB 591 ATDQGTGVVDASSLTNAQTTLTINGKI---GTIGANNK-----TLGQ-FNIGS 632

Query Match 3.0%; Score 199.5; DB 1; Length 1654;
Best Local Similarity 19.9%; Pred. No. 0.018;
Matches 273; Conservative 178; Mismatches 584; Indels 337; Gaps 63;
QY 15 SLVNSAVSGFDTTTLKNDQYQALNGQLVLLR--FSKAADGNPSGDNVTYDQFSKNYAT 72
DB 337 SVDNGKATIDGGVYAKDMVYQSANATGQVNFRIYDVGDGT-TAFKTAASKVTITQDS 395
QY 73 TGGNFDY-----VKVNGNQVFESGHWATNQSKDKDSQWIIVLVNGKEVRQLVNDTKEG 136
DB 396 MFGNTDFGNLAAQIKV-PNALTGTGNETGDSNPGTAGVITFD-----ANGLES 445
QY 127 AAGFNRDVYKVPAINENS-----SMSCFQGIITLPVTVKXENVOLVHRFSND---VKTGE 179
DB 446 ASA-----DANVAVTNNTATETASG-AGVVQLSGTHAAE-----LRLGNAGSIFKLAD 492
QY 180 GNYVDFWSELMPVKDSFKGNGPLKQ-----TGLQTINGQOYYIDP----- 220
DB 493 GTVI-----NGKVNQALVGGALAAATITLTDGSAITIDGIGNAGGAAL 536
QY 221 ---TTGQPRKNFLIQSGNNWYIFDSDTGVGTNALELQFAKGTVSSNQYRN---GNAAY 273
DB 537 QRITLANDAKKTLTLGKANII-----GAGGCIIDLQANGGTIKLSTQNNIVYVDFDLAI 590
QY 274 SYDDKSIENYNGYLTADTYWRPKQILKDGTTWTSKEDMRPILMVMWPNLTQAYVLYN 333
DB 591 ATDQGTGVVDASSLTNAQTTLTINGKI---GTIGANNK-----TLGQ-FNIGS 632

Matches 272; Conservative 175; Mismatches 533; Indels 496; Gaps 70;

```
QY 12 VPSLYNSAVSGEDTTIKLNDQOALNGQLVLRFSKAADGNPSGNTYTD-----64
Db 11 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
316 VFTNSDTSATTSATLSAIEDAKIGADVLNMSL-----GSDSGNOTLEDPELAAYQ 368
QY 65 -----QFSKNVATGGNEDYVYKNGNQVEFSGNHATN-----QSNQD 100
Db 11 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
369 NANESGTAATVTSAGNSGTSGGATGEGVNDYTGILQDN--EMVGTGTSRGATVASAENTD 426
QY 101 KDSQWIIY-----LVNGKEVRKQLVNDTKEGAAGFNMRNDYKVNPAIENSS-----MS 148
Db 11 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
427 VITQAVTITDGTGLGLGPETICLSSNDP--TGSEPDQKFFVYVVDASGNLSKGVADYTA 483
QY 149 GEQGIITLPTVYKNEVOLVHRFSNDYKTVGSGNYVDFSELMVVKDSQKNG-----201
Db 11 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
484 DAKGKIAI-----VKRGELTFAD-----KQKYAQAAGAAGLIIV 517
QY 202 -----PLKQFGLQTLINGQQYYIDPTVQCPKKNFLQSGNNWYFDSGTG 245
Db 11 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
518 NNDGTATPVTSMALTTFTFGLSVTGCKLV-----DWAHAHPDSS 559
QY 246 VGTN-ALBQFAKCTVSSNEQYRGNAAYSDDKSIEVNGYLTDWYRPKQILKDGTT 304
Db 11 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
560 LGVKIAL-----TLVPNQYV-----EDKMSDFTSYGVPYNSLSFKDITAPGNI 604
QY 305 W-----TDSKETDMRPILVWVWPVILQA-----YYLNTMKOHGNLLPSALPFFN 349
Db 11 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
605 WSTONNNGYTNMSTGSMASPIFAGSQALLKQALKNKNPFIYAYYKQLKGTALTDFLKTYE 664
QY 350 AD-ADPAELNHYSEIVQCNLEKRISETGNTDWTLMHDFVTNNPMMNKDSNVNFSGIQ 408
Db 11 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
665 MNTAQPIINDIYNVNI--VSPRQAGLVDVKAAL--DALEKNP-----STVV 708
QY 409 FQGGFLKYVNDLTPYANSVRLGLRMPINTKDQYRGEQFELLANDIDNSPVVQAEGLN 468
Db 11 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
709 AENGYPAVELXDFTS-TBKTEKL-----TFETNETHELTQYMDSNTDTNAVYVSATDPN 761
QY 469 --WLYLLNFGTITANNDQAFNDSVRYDAPDNIDALMNIQAQDFNAAV-----515
Db 11 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
762 SGVLV-----DKKIDGAIAIKAGSNITVPAGKTAQIEFTLSLPKSFQQQPFVE 808
QY 516 -----GMSDAVSXNKHILEDNWHD--PEYFNKI-----GN-----PQLTMDDTIKNS 558
Db 11 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
809 GFLNPKSGDGSRLNLPYMGFGDNDGKIVDSLNGIYSPAGGFGVPLTNKNIGTOY 868
QY 559 LNHGUSDATNKGDLDAIVHQSLA--DRENNSTENVYPIYSFVRAHDNNSQDQIQNAIDV 617
Db 11 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
869 YGGWVTDAD--GNQVDDQAIAPFSSDKNALYNDISNRYLLRNISN-----VQVDILDG 920
QY 618 TGDYHHTFEDEQGDIDAYIODONSTVKYNLNPASAYAILLTNKTIPRVYGGDLYT 677
Db 11 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
921 QGNKVYTLN-----SSINRKYTYNAHSQQYI-----YIHAPAW 954
QY 678 DGGQYMEHQTRYDPTLNLLKS-----RVKYV-AGGQSQMOTMSVGGN--NNILTSVRY 727
Db 11 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
955 DG-----FYIDQDGNIKTADGSGYTYRISGVPEGDKRQVDFPFKLSKAPTVRH 1006
QY 728 CKGAMTADTGTDETRQ-----GIGVVVSNTPNLKLGVNDKVVYLHMGAAHN 775
Db 11 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1007 -----VALSAKTENGKTQYLTABAKDLSGLDA-----TKSVKTAINE--VTNLDATFTD 1055
QY 776 QYRAAVLTTTGDGVNNT-----SDQAPVAMTDENGDLVLSHNLVYNGKEADTAV 828
Db 11 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1056 AG-----TTADG---YTKIETPLSDEQAQALGNDGNSAEIYLTDN---ASNATDQDASV 1103
QY 829 QYANPDYSGYLVAWVPVGSADNDQATAPSTERNKSGNSAVRTNAAF-----875
Db 11 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1104 Q---KPGSTSFDLIVNGGIPDKLSSITTTGAEANTQGGTTFYFSGTYPAAVDSYTTDAQG 1160
QY 876 ---DSWTFEAFNSFYPTTKESERANVRIAQNAADFASLGFTS-----FE-----MAP-- 921
Db 11 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1161 KKHDLNLTYYDATN---SFTVASMPVTNADYAAQVDLYADKAHTQLLKHFDTKVRLTAPTE 1217
```

```
QY 922 ---QYNSKSDRTFLDSTI-----DNGYAFETDRYDLQWSBPNKYG-----957
Db 1218 TDLKFNNGSDQT-SEATIKVTGIVSADTKTVNNGDVTVAALDAQHHSVDVPPVYVGDNTIK 1276
QY 958 ---TDEDLRNAIQALHAGLQVADWVDPQIYNL-----PGK 991
Db 1277 VTATDEGNTTTEO-----KTIITSSYDPMKLNKSVTFDQGVTFGANEFNATSAKFVDPKT 1331
QY 992 EVATVTT-----RVDDRGNNVWKDAIINNLLYVNTIGGGYQKKYGSFIDKLQ-K 1040
Db 1332 GIAVITGVKVRHPTTFLQVDGKQIPKIDDLFSFLDGLCTLG---QKPFQVYVGDITTONK 1387
QY 1041 LYPE--IPTKKQVSTGVAIDPDSQKITEMSACYFNGTNHLHRSYGVYLVKADGGQYNNL---1095
Db 1388 TFQALATFILDVAPTILSDSSTD---APVYTNPNFQITGTA---TDNAQYLSLSIN 1439
QY 1096 -GTTTKQFLPQLTGEKKQGNFVKNDGNGYFYDLAGNNVKNFTFIEDSVGNWYFFDQD 1154
Db 1440 GSSVASQYVDINISNGK-----PGHMA-----IDQP 1465
QY 1155 GKMYENKHFVDV---DSYGEKGTFFELKNGYSFRGLVQTDNGTYYPDNYGKMYRNOTIN 1211
Db 1466 VKLLEGKNLTVAVTDS--EONT-----TTKNITYVYE-----PKKTLA 1502
QY 1212 AGAMIYILDENGKLKASYNDSABYPT---STDVGK 1244
Db 1503 APTVTPSTTEPAKTVTTLTANSAAATGETVQVYSADGGK 1538

RESULT 19
OMPE RICPR STANDARD; PET; 1643 AA.
ID OMPB-RICPR Q53020; Q9ZCM0;
AC 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Outer membrane protein B precursor (188 kDa surface-layer protein)
DE (Surface protein antigen) (cell surface antigen 5) (Scas5) (rOmpB)
DE (rOmp B) [Contains: 120 kDa surface-exposed protein (surface protein
antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide].
GS OMPB OR SPAP OR SPA OR RP704.
QN Rickettsia prowazekii
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=782;
[1]
SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN-BREINL;
RX MEDLINE=91045972; PubMed=2122457;
RA Carl M., Dobson M.E., Ching W.M., Dasch G.A.;
RT "Characterization of the gene encoding the protective paracrystalline-
surface-layer protein of Rickettsia prowazekii: presence of a
truncated identical homolog in Rickettsia typhi";
RL Proc. Natl. Acad. Sci. U.S.A. 87:8237-8241(1990).
RN [2]
SEQUENCE FROM N.A.
RC STRAIN-BREINL;
RA Moron C.G., Yu X.J., Walker D.H.;
RT "Sequence analysis of ompB of Rickettsia prowazekii.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
SEQUENCE FROM N.A.
RC STRAIN-MADRID E;
RX MEDLINE=99039499; PubMed=9823893;
RA Andersson S.G.E., Zomrodipour A., Andersson J.O.,
RA Slicheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT "The genome sequence of Rickettsia prowazekii and the origin of
mitochondria.";
RL Nature 396:133-140(1998).
RN [4]
PARTIAL SEQUENCE.
```


QY 1087 ADGGQYVNLGTTTKQFLPI-QLTGEKKQG--NEGFKVGN--DGNVYFYVDLAGN--MKVNT 1139
 Db 1386 TT-GVVVGLDTLASDNLMIGAATGTTKDIKHQDYKKGDKTDINGLSFSLYGSQQLVKNF 1444
 QY 1140 FTE-----DSVGNWYFFDDGKKVKNKHFVDVDSYGEKGYFFFLKNGVSRFGGLV 1189
 Db 1445 FAQGNALFTLKNVKSORSYFFESNGKSKQIARAGNYDNMFGGNLIPGYDYNAMPNVLV 1504
 QY 1190 QTDNGTYTF---DNY---GKMVRNQTIN-----AGAMIYTIID----- 1220
 Db 1505 TPMAGLSYLKSSNENYKETGTTVANKRINSKESDRVDLVGAKVAGSVTNITDIYIPEI 1564
 QY 1221 -----ENGL-----IKASYN-----SDA--EPTSTD 1241
 Db 1565 HSFVHVKNYNGKLSNSQSMLDGQTAPFTSQPDRTAKTNYIGLSANIKSDAKMEYIGYD 1623

RESULT 20

OMP_RICTY STANDARD; PRF; 1645 AA.
 AC P96989;
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE Outer membrane protein B precursor (168 kDa surface-layer protein)
 DE (Surface protein antigen) (Cell surface antigen 5) (Scat) (rOmpB)
 DE (rOmp B) [Contains: 120 kDa surface-exposed protein (Surface protein antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide].
 GN OMPB OR SLF.
 OS Rickettsia typhi.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OC Rickettsiaceae; Rickettsiae; Rickettsia.
 OX NCBI_Taxid=785;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-WILMINGTON;
 RX MEDLINE=94040787; PubMed=8224886;
 RA Hahn M.-J., Kim K.-K., Kim I., Chang W.-H.;
 RT "Cloning and sequence analysis of the gene encoding the crystalline
 surface layer protein of Rickettsia typhi.";
 RL Gene 133:129-133(1993).
 RN [2]
 RP PARTIAL SEQUENCE.
 RC STRAIN-WILMINGTON;
 RX MEDLINE=92114896; PubMed=1370573;
 RA Ching W.M., Carl M., Dasch G.A.;
 RT "Mapping of monoclonal antibody binding sites on CNBr fragments of
 the S-layer protein antigens of Rickettsia typhi and Rickettsia
 prowazekii.";
 RL Mol. Immunol. 29:95-105(1992).
 RN [3]
 RP IDENTIFICATION OF CLEAVAGE SITE.
 RA MEDLINE=92104668; PubMed=1729180;
 RA Hackstadt T., Nesser R., Cleplak W., Peacock M.G.;
 RT "Evidence for proteolytic cleavage of the 120-kilodalton outer
 membrane protein of rickettsiae: identification of an avirulent
 mutant deficient in processing.";
 RL Infect. Immun. 60:159-165(1992).
 CC -1- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR
 STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
 VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.
 CC -1- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR.
 CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A
 S-LAYER WITH HEXAGONAL SYMMETRY.
 CC -1- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPB FAMILY.
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

CC or send an email to license@sib-sib.ch.

DR EMBL; LO4661; AAB48987.1; -
 DR InterPro; IPR003858; rOmpA_rOmpB.
 DR Pfam; PF02708; rOmpA_rOmpB; 1.
 KW Antigen; S-layer; Transmembrane; Cell wall.
 FT CHAIN 1 1353 120 KDA SURFACE-EXPOSED PROTEIN.
 FT CHAIN 1354 1645 32 KDA BETA PEPTIDE.
 FT TRANSHEM 1415 1429 MEMBRANE ANCHOR (POTENTIAL).
 FT CONFLICT 657 657 H -> N (IN REF. 2).
 FT CONFLICT 842 842 V -> I (IN REF. 2).
 FT CONFLICT 1071 1071 G -> A (IN REF. 2).
 FT CONFLICT 1306 1306 G -> S (IN REF. 2).
 SQ SEQUENCE 1645 AA; 169698 MW; 0CB5641C7EB185EE CRC64;

Query Match 3.0%; Score 198; DB 1; Length 1645;

Best Local Similarity 19.3%; Pred. No. 0.021;

Matches 318; Conservative 194; Mismatches 592; Indels 542; Gaps 79;

QY 22 SGFDTT---IKL---TND-----QYCALNGQIQVLLRFSKAADGNPSGDNVTDQ 65
 Db 53 AGFQDTGAGVNLPAVATNSVITANSNNAITFTPNGNLNSLFLDTANTLAVTINENT-TLG 111
 QY 56 FSKNYATGGNFDYKVGNGQVESHGHAHQ--SNDKDSQWIIIVLNGKEVKRQLVNDT 123
 Db 112 FVTNVTGKGNFFNETIGAGKSLTITGHGITAQAARTKSAQNVVSKVN----- 159
 QY 124 KEGAGGNRNDVYKVNPAIENSSMSGFGQITLFPVVK-----NEVQLVHRFSNOVKTG 178
 Db 160 --AGAAINDLSGVSIDTAAPSVEFLNLTTPQAPLTLGDNAKIVN-----G 209
 QY 179 EGNVYDFWSELMPVKDSFQKNGPLKQF-GLQTN--GOQYIIDPTTGPQRKNFLQSGN 235
 Db 210 ANGLINITNGFVKVSD-----KTFAGIKTINIGDQGLMFTTTPDAANALNLAGGG 260
 QY 236 NWIYFDSOTGVGT-----NALELQFA-----KG-----TVSSNEQYRNG---NA 271
 Db 261 NTINFGSDGTGKLVLVSKNGNAFEFNVTSGLGNLKGVIIEFTDTAAAGKLIANGGAANA 320
 QY 272 AYSYDD-----KSIENVNGYLTADTWYRPKQILKDGTTW-----T 306
 Db 321 VIGTDNGAGRAAGFIVSDNGNATISGOYAKDIVLOSANAGQVTFEHLVDVGLGGKT 380
 QY 307 DSKETDMRPIL-----MWMPTLTPQAYLYLNTYKHOGHLLPSALPFFN 349
 Db 381 NFKTADSKVITENASFGSTDFGNLAVQIVVPNNKILTNFGIGDAKNNGN--TAGVITPN 438
 QY 350 ADADPAELNHYSVIVQONI-----EKRISETGNTDMLR--TLMHDF 388
 Db 439 ANGLVSGNTDPIVVTNIXAIEVEGAGIVOLSGIHGAELRIGNAGSIFKLADGTVINGP 498
 QY 389 VTNNPMWNKDSENVNFGIOFGGFLKYENSDLTPPYANSYRLILGRMPINIKQDTYRGQE 448
 Db 499 VNQNPLVNNNA--LAAGSIQLDGSAI-----ITGD--INGAVNA-----ALQD 538
 QY 449 FLIAND-----IDNSNPVYQABQLWLYLLNFGTI-----TANNOQANFD-SVYVDAPD 497
 Db 539 ITLANDASKILTISGAN-IIGANAGGAIHFQANGGTITLTSTQTNILVDFDLDTDTQGT 597
 QY 498 NIDA-DLMTIAQDYFNAA---YGMDSQAVSNKHINILEDWNHADPEYFNKI-----GNPQ 548
 Db 598 VVDASLTNQTLTINGSICTIGANYKYLQRFVVGSKTTLNAGDVAINELVWENQDSVH 657
 QY 549 LT-----MDPTIKNSLNHG-----LSQATNRWGLDAIVHOSLADRENNSTENVVPIYSF- 598
 Db 658 LTHNTYLTITKTI-NAANQGIIVAADPIN-----TDTALADGTNLGSAESPLSNHFA 709
 QY 599 VRAHDNNSQDQIQWAI-----RDVTGDKYHTFTTEDEKQ---TDAYIQDNSTVKKYNLY 651
 Db 710 TKRANGDSILHIGKGNLYANNITTTDANVGLSHFRSGGTSIVSGTVGGQOG-LKLNNL- 767
 QY 652 NIPASVAILLTNKTIPRYYVDLYTDGGQYMEHQ-----TRY----- 690

Db 768 -----ILNGTIVK--FLGDTFNGTIEGKSILOISSNYITDHIIESADNIGTLEF 817
Qy 691 ---DTLTLMLKSRVYVAGGSMQTMVGGNNILTSVRYGKGAMTATDTGTDTRQGI 747
Db 818 VNTDPIVTLNKGQAYF--GVYKQVMVSGPNIAFNEIGNVGVAHIAVDSISFENASLGA 875
Qy 748 GV-VVSNTP-----
Db 876 SLFLISGTPDLVLAIKSTGVGTVDNFNAPILVYSGIDSMINNGVIGDQKNIIALSIGS 935
Qy 762 NDKVVLH---MGAHKNKQYRAAVLTITDG-
Db 936 DNSITVNSNLIAGIRTTKQGVFTLSSGIPNPNPTIYGLENGDKLQKVTFITDIN 995
Qy 798 -----APVATDENGDLXLSHNLVYNGK
Db 996 MLGSIATNVTINDVTLTGGIAGTDFDGTGILTSINGNANVFPVDRFESHPTSMIVST 1055
Qy 831 YANPDVSGYL--AVWVPVGSAD-----
Db 1056 KANQGVTVYLGALVGNIGSSDIPVASVRFGTGNDGSGVLQNIHSQNIIDFGTYNLTILNS 1115
Qy 865 ---GNSAYRTNAAFD---SNVIFEAFSNFYVTTKESERANVR-----IAO----- 904
Db 1116 DVILGGGTATNGEIDLTLNLI FANGSTGWNTSLSTTLNVSGNVGQIVIAEGAOVN 1175
Qy 905 -----NADFFASLQFTSEMAPQYNSSKDRFTFLDSTIDN----- 938
Db 1176 ATTGQTTPIKIQDANANFSGTQVYTLQGGARENGTLGAPNFDVGTGNIFVKVELYRDA 1235
Qy 939 --GVAFTDRYDL-----GMSEPNKYGTDEDLRNAIQALHKAGLOVM-----A 978
Db 1236 NQDYLVTFTNDVLNVVTVAVGNSALANAPGVHQAICLESTDTAAAXNNMLLAKDSSDAV 1295
Qy 979 DWVPDQIYNLPKKEVATVTRVDRGNVWKDAIINNLL-----YVVTI-----ICGGEY-QKK 1029
Db 1296 TFI-GAATDGTGAATVATVNDTQKT---QDLGNRLGALRYLSNSETADVGGSETGAVS 1351
Qy 1030 YGGAPFLKQKLEIFEITFKKQVSGVAIDPSQKTESAKYFNGINILHRG--SGYVLKA 1087
Db 1352 SGDEAID-----QVSGYV-----WAKPFYNIAEQDKGGLAGYKRAKT 1388
Qy 1088 DGGQVYNLGTGTTKQFLPQLA-----GEKKQNGEFGVKGNGCNVYFYDLAGNMVK 1137
Db 1389 --AGVVVGLDTLANDNLMIAGAIGTKTDIKHQDYKKGDKTDIKGLSFLSG---AQQLVK 1444
Qy 1138 NTFIDSV-----GNVYFFDQDKMYENKHFV--DVDSYGKGTYPFLKNGYSFR 1185
Db 1445 NFFAQGSAIFTLNKKVKSQRYFFDANGKM--NQIAGNVNDNITFGNLMFGYDYNALQ 1502
Qy 1186 GGLVQTDNGTYF---DNY---GKMYRNOTINA-----GA-----MI 1216
Db 1503 GVLVTPMAGLSYLSKSSNENYKGTGTVANKRIHSKFSRDLIVGAKVGTGSAMNINDIVI 1562
Qy 1217 Y-----TLDENGKLKASYNDAE 1235
Db 1563 YPEIHSFVHVHVNGKLSAQSMLDGQ 1588

RESULT 21

120K_RICRI

ID 120K_RICRI

AC P14914;

DT 01-APR-1990 (Rel. 14, Created)

DT 01-FEB-1996 (Rel. 33, Last sequence update)

DT 01-MAR-2002 (Rel. 41, Last annotation update)

DE 120 kDa surface-exposed protein.

GN P120.

OS Rickettsia rickettsii.

OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;

CC Rickettsiaceae; Rickettsiidae; Rickettsia.

OX NCBI_TaxID=783;

RN SEQUENCE FROM N.A.
RP STRAIN-R;
RC MEDLINE-90136087; PubMed-2515418;
RA Gilmore R.D. Jr., Joste N., McDonald G.A.;
RT Cloning, expression and sequence analysis of the gene encoding the
RL 120 kDa surface-exposed protein of Rickettsia rickettsii,"
RM Mol. Microbiol. 3:1579-1586(1989).
CC -1- RICKETTSIAL VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.
CC -1- FUNCTION: MAJOR STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS
CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A
CC -1- S-LAYER WITH HEXAGONAL SYMMETRY.
CC -1- MISCELLANEOUS: ITS C-TERMINUS POTENTIALLY MAY BEAR THE EPITOPES
CC -1- CONFERRING ANTIGENICITY TO THE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPB FAMILY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).

EMBL: X16353; CAA34402.1; --
PIR: S07575; S07575.
DR InterPro; IPR003858; rOmpA_rOmpB.
DR Pfam; PF02708; rOmpA_rOmpB; 1.
KW Antigen; Glycoprotein; Cell wall; S-layer.
FT CARBOHYD 7 7 N-LINKED (POTENTIAL).
FT CARBOHYD 66 66 N-LINKED (POTENTIAL).
FT CARBOHYD 86 86 N-LINKED (POTENTIAL).
FT CARBOHYD 103 103 N-LINKED (POTENTIAL).
FT CARBOHYD 147 147 N-LINKED (POTENTIAL).
FT CARBOHYD 268 268 N-LINKED (POTENTIAL).
FT CARBOHYD 330 330 N-LINKED (POTENTIAL).
FT CARBOHYD 375 375 N-LINKED (POTENTIAL).
FT CARBOHYD 415 415 N-LINKED (POTENTIAL).
FT CARBOHYD 424 424 N-LINKED (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (POTENTIAL).
FT CARBOHYD 444 444 N-LINKED (POTENTIAL).
FT CARBOHYD 515 515 N-LINKED (POTENTIAL).
FT CARBOHYD 547 547 N-LINKED (POTENTIAL).
FT CARBOHYD 593 593 N-LINKED (POTENTIAL).
FT CARBOHYD 655 655 N-LINKED (POTENTIAL).
FT CARBOHYD 698 698 N-LINKED (POTENTIAL).
FT CARBOHYD 710 710 N-LINKED (POTENTIAL).
FT CARBOHYD 799 799 N-LINKED (POTENTIAL).
FT CARBOHYD 800 800 N-LINKED (POTENTIAL).
FT CARBOHYD 826 826 N-LINKED (POTENTIAL).
FT CARBOHYD 844 844 N-LINKED (POTENTIAL).
FT CARBOHYD 861 861 N-LINKED (POTENTIAL).
FT CARBOHYD 879 879 N-LINKED (POTENTIAL).
FT CARBOHYD 920 920 N-LINKED (POTENTIAL).
FT CARBOHYD 926 926 N-LINKED (POTENTIAL).
FT CARBOHYD 1116 1116 N-LINKED (POTENTIAL).
FT CARBOHYD 1128 1128 N-LINKED (POTENTIAL).
FT CARBOHYD 1140 1140 N-LINKED (POTENTIAL).
FT CARBOHYD 1146 1146 N-LINKED (POTENTIAL).
FT CARBOHYD 1211 1211 N-LINKED (POTENTIAL).
SQ SEQUENCE 1300 AA; 132801 MW; E09E52C3F647243D CRC64;

Query Match

Best Local Similarity 19.6%; Pred. No. 0.021;

Matches 258; Conservative 169; Mismatches 536; Indels 354; Gaps 59;

Qy 39 NGQLQVLLRFSKAADGNPSGDNVTVDQFSKNYATITGGNFYKVNG-NQVEFSGWHATNQ 97

Db 86 NGTLE-----SASADANVAVTNIT-----AIEASGAGVQLSSTHAA-- 123

Qy 98 SNDKDSOIWILVNGKEVKRQLVNDTKEGAG--FNRNDYKVNPAIENSSMSG---FQG 152

Db 124 -----ELRLNAGSIFKLAQGVINGKYNQALVGGALAAAG 159
QY 153 IITLPVTVKNENQVLVHRFSNDVATGSGNVYDFWSELMPVKDFQKGNGLPKQFGLQTIN 212
Db 160 TITLDGSA-----TITGDIGNAG-----AAALQRI- 185
QY 213 GQQYIIPDTGPQRKNELLASGNWIIYFSDTGVGTWALELOFAKGVVSSNEQYRN----- 268
Db 186 -----TLANDAKKTITLGGANII-----GAGGCTIDQANGSTIKLTSTQNNITVVD 231
QY 269 GNAAYSDDKSIENNVGLVADTWYRKPQILKDGCTTWTDSKETDMPRLVWVWPTLTOA 328
Db 232 FDLAIVDQTVGDVASSLTAQTITNGKI-----GTIGANNK-----TIGQ- 273
QY 329 YLLNMYK-----QHGNLLPSALPFFNADAPALNHSYSEIVQONIEKRISERGNTDMLRTLM 385
Db 274 FNIGSSKTVLSNGVAINELVIGNDGA--VQFAHDTYL-----ITRTNAAGQGGKII 323
QY 386 HDTVTN-----NPM--WNKDSNVNFSGIQPOG--GFLKYENSULTPYANS 427
Db 324 FNPVYNGTTLAAGTNLGSATNPLAEINFGSKGVNVTVLNVGSGVNLXATNITTTDANV 383
QY 428 DYRLILGRMPINIKQOTYRGQB-----FLLANDIDNSNPVQAE--OLNW 469
Db 384 GSFVENAGTNIVSGTVGGQGNKFNVALENQTVKFLGNATFNGNTTAA NSTLQIGG 443
QY 470 LY-----YLLNFG--TITANNDOANFDSVR--VDAPONIDADLMIADQY 510
Db 444 NYTADCVASADGTGIVEFVNTGPITVTLNKAQAPVNAKQITVSGPGVNVINEIGNAGNH 503
QY 511 FNAAYGWDSDAVSNKHINILEDWHADEPFENKIGNPQ--LTHDDTINKSLNHGLS--DATN 568
Db 504 HGAV--TDTAFENSSGLGAVVFLRGP--FNDAGNTMPLTIKSTVGNTAKGPDVPSVV 559
QY 569 RWGLDAIV--HQSADREN-----NSTENVIPNYSFVRA-----HNNNSQDQIQNAIRD 616
Db 560 VLGVDVIAQGVIGDQNNIVGLGSDNGIIVNATLYAGISLNNNQGVTVLSGGVPH 619
QY 617 VTGKDYHTTFEDBQKIDAYIQDQNSTVKKY--NLYNIPASYAIL-----LTNKDITPRVY 671
Db 620 TPGVYVYGLT-----GITAGKFRQVFTTDDYNLGNLIATNATINDGVTVTTGGIAGIG 673
QY 672 YGDLTYDGGQYNEHQVRYDPLTLNLLSKRYKVVAGSQSMQTM-----SVGGNNNLTLS 724
Db 674 FDGKITLGSVNGNWRFAADGILSNSTSMIGTTKANKGTVTYVLGNAPVGNIGSDTPVAS 733
QY 725 VRYKGAMTATDTGDETRTQIGVYVSNIPNLKLVNDKYVLMHGAHKNOQYRAA--V 782
Db 734 VRE-----TGSDSGAGLQGNYSQVIDFGTVNLGI--VNSNILLGGGTTAINGKIDLYTNT 787
QY 783 LTTVDVLYNTSDGAPVAMTDENGDLVLSHNLVNVNGKEADTA-----VQGYANPD 835
Db 788 LTFASGTSTWGNNTSIETITFLANGNT---GHIVILEGAQVNTTGTITIKVQDNANAN 844
QY 836 VSGYLAVVVPVGSADNQDARTAPTEKNSGNSAVRTNAPDSNVIFAFSFFVYTPPKES 895
Db 845 FSGVQTYLIGGARFNGLTGSFNF-----AVTGSNRFVYSLIRANQDYVITRNNNA 898
QY 896 ERANVRIAQNADFFASLG-----FTSFEMAPQ-----YNSKSDRTFLDSTIDNGYA 941
Db 899 ENVVNTDIANSPPFGAPGVQDQNVTTFNATNTAAYNLLAKNSANSANFVGAIVTDTSA 958
QY 942 FTDRLVLCMSEPKYGTDEDLRAIQAHLKAGLOVMAWDPQIYNLPGKEVATVTRVDD 1001
Db 959 AITNVQLDLAK-----DIQAQGLNRLGALRYLGTPEAE-----NAGPEFAGISAAYA 1006
QY 1002 RGNVWKDAIINNLYVWNTIGGGYQKYGAGFLDKLQKLYPEFTTKQVSTGVAID--- 1058
Db 1007 AG-----DEADINVAI-----GIWAKPF-----YDAHQSKKGLAGYAKTGVVIGLDT 1052
QY 1059 -PSQKITEWSAKYFNGTNHLRGSGYVLKADGGQYVNLGTTTKQFLPQLTGTGKKNQNEG 1117

Db 1053 LANDNEMIGAAGITKTDIKH-----QDYKKGDKT----- 1082
QY 1118 FVKGNDQNYFYFDLAG--NMVKNTFTIEDSV-----GNWYFFDQDG---KMYENKH 1162
Db 1083 -----DVNGFSFSLYGAQQLVKNKQRYFFDANGNNSKQIAAGH 1137
QY 1163 FVDVDSVGEKGTTFEFLKNGVSFRGGLVQTDNGYIF-----DNY-----GKMYRQNTINA 1212
Db 1138 Y-DNMTFGGGLTGVGYDYN--AMOGVLVTPMAGLSYLKSSDENYKGTGTTVANKQVNS 1191
RESULT 22
IGAL_HAEIN
ID IGAL_HAEIN STANDARD: PRT; 1541 AA.
AC P42782;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DE 01-NOV-1995 (Rel. 32, Last annotation update)
DE Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGAL protease).
GN IgA.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OC NCBI_faxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HK368 / SEROTYPE B;
RX MEDLINE=89379374; PubMed=2506130;
RA Poulsen K., Brandt J., Hjorth J.P., Thøgersen H.C., Kilian M.;
RT *Cloning and sequencing of the immunoglobulin A1 protease gene (iga)
of Haemophilus influenzae serotype b.;
RL Infect. Immun. 57:3097-3105(1989).
RN [2]
RP MOTAGENESIS OF SER-288.
RC STRAIN=HK368 / SEROTYPE B;
RX MEDLINE=92234949; PubMed=1373717;
RA Poulsen K., Reinholdt J., Kilian M.;
RT *A comparative genetic study of serologically distinct Haemophilus
influenzae type 1 immunoglobulin A1 proteases.*;
J. Bacteriol. 174:2913-2921(1992).
CC -I- FUNCTION: VIRULENCE FACTOR; CLEAVES HOST IMMUNOGLOBULIN A
PRODUCING INTACT FC AND FAB FRAGMENTS.
CC -I- CATALYTIC ACTIVITY: Cleavage of immunoglobulin a molecules at
certain Pro-I-xaa bonds in the hinge region. No small molecule
substrates are known.
CC -I- SUBCELLULAR LOCATION: Secreted.
CC -I- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC
SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE
DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY
SIMILARITY).
CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6 (SERINE PROTEASE).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed, usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC EMBL; X64357; CAA45708.1; -
CC EMBL; M87492; AAA24969.1; -
CC MEROPS; S06.001; -
CC InterPro; IPR000710; Iga_S6.
CC Pfam; PF02395; IGAL; 1.
CC PRINTS; PR00921; IGASERPTASE.
KW Hydrolase; Serine protease; Transmembrane; Zymogen; Signal.
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 1008 IMMUNOGLOBULIN A1 PROTEASE.
FT PROPEP 1009 1541 HELPER PEPTIDE (POTENTIAL).
FT ACT_SITE 288 288 PROBABLE.
FT MUTAGEN 288 288 S->T: LOSS OF ACTIVITY.

TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-812-829-2

Query Match 5.8%; Score 165.5; DB 3; Length 750;
Best Local Similarity 20.3%; Pred. No. 8.1e-06;
Matches 96; Conservative 62; Mismatches 160; Indels 155; Gaps 20;

QY 7 EFLANDIDNSNPVQAEOLNWLNYLLNFGTITANDQANFDSVRVDAPDNIDADLNMIA 66
DB 223 DYLWADIDYDHPDVAEKRW-----GTWYAN--ELQIDGNRLDAV----- 262
QY 67 QDYFNAYGMDSDAVSNKHI--NILEDW-NHADPEYFNKIGNPOLTMDDTIKNSL----- 118
DB 263 -----KHIFSFURDWNHVR-----EKTCKEMFTVAEYWQNDLGALEN 301
QY 119 -----NHGLSDATNRGLDAIVHQSLADRENNSTENVIPNYSFVRAHDNNSQDQIQN 171
DB 302 YLNKTNFHNSFVDFPLHYQFHA-----ASTQGGGYDMRKLNGTVVSKH----- 345
QY 172 AIRDTGKDYHTTFEDQKGDYAIQDNSTVKYKLYNIPASVAILLTNKTIPRVY 231
DB 346 PLKSVTFVDNH-----DTQPG-----QSLESTVQTWFK---PLAYAFILTRSGYPQVY 392
QY 232 GDLYTDGQYMEHQTRYDYDTLNLKSRVKYVAGGQSMQTSVGGNNILTSVRYGKGM 291
DB 393 GDMYGTGDSQREIPALKHKIEPILKARKQYAYGAQH---DYFDHHDIVGWTREGDSSV 448
QY 292 TATDTGTDTRTQGGIGVVVSTPNLKLGVNDKVLHMGAAHKNQYRAAVLTITDGVINY 351
DB 449 A-----NSGLAALITDGP-----GGAKR-----MY 468
QY 352 TSDQAPVAMTDENDLYLSSHNLVNGKEADTAVQGYANPDVSGYLAVWVP----- 404
DB 469 VCRQNACTWHDITCN---RSEFPVINSEGWGEFHVNG-----GSVSIYVQSGGPGT 518
QY 405 -----VGASNDQARTAPSTKESNAYTNAAFDANSVIFEAFSNVFYPT 451
DB 519 PNNGRGIGYIENGNTVYSNIDFGSGATGFSATVATEVNTSIQIRSD---SPT 568

RESULT 9
US-08-814-052-6
Sequence 6, Application US/08814052
Patent No. 6015783

GENERAL INFORMATION:
APPLICANT: von der Osten, Claus
APPLICANT: Cherry, Joel R.
APPLICANT: Bjornvad, Mads E.
APPLICANT: Vind, Jesper
APPLICANT: Rasmussen, Michael Dolberg
TITLE OF INVENTION: PROCESS FOR REMOVAL OR BLEACHING OF SOILING
TITLE OF INVENTION: OR STAINS FROM CELLULOSIC FABRIC
NUMBER OF SEQUENCES: 55
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6015783o No. 6015783disk of No. 6015783th America, Inc.
STREET: 405 Lexington Avenue, Suite 6400
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/814.052
FILING DATE: 06-NAR-1997
CLASSIFICATION: 510

ATTORNEY/AGENT INFORMATION:
NAME: Iamdiris, Elias J
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 4684.204-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
TELEX:
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 560 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-814-052-6

Query Match 5.5%; Score 158; DB 3; Length 560;
Best Local Similarity 20.8%; Pred. No. 2.3e-05;
Matches 88; Conservative 62; Mismatches 155; Indels 122; Gaps 19;

QY 7 EFLANDIDNSNPVQAEOLNWLNYLLNFGTITANDQANFDSVRVDAPDNIDADLNMIA 66
DB 223 DYLWADIDYDHPDVAEKRW-----GTWYAN--ELQIDGNRLDAV----- 262
QY 67 QDYFNAYGMDSDAVSNKHI--NILEDW-NHADPEYFNKIGNPOLTMDDTIKNSL----- 118
DB 263 -----KHIFSFURDWNHVR-----EKTCKEMFTVAEYWQNDLGALEN 301
QY 119 -----NHGLSDATNRGLDAIVHQSLADRENNSTENVIPNYSFVRAHDNNSQDQIQN 171
DB 302 YLNKTNFHNSFVDFPLHYQFHA-----ASTQGGGYDMRKLNGTVVSKH----- 345
QY 172 AIRDTGKDYHTTFEDQKGDYAIQDNSTVKYKLYNIPASVAILLTNKTIPRVY 231
DB 346 PLKSVTFVDNH-----DTQPG-----QSLESTVQTWFK---PLAYAFILTRSGYPQVY 392
QY 232 GDLYTDGQYMEHQTRYDYDTLNLKSRVKYVAGGQSMQTSVGGNNILTSVRYGKGM 291
DB 393 GDMYGTGDSQREIPALKHKIEPILKARKQYAYGAQH---DYFDHHDIVGWTREGDSSV 448
QY 292 TATDTGTDTRTQGGIGVVVSTPNLKLGVNDKVLHMGAAHKNQYRAAVLTITDGVINY 351
DB 449 A-----NSGLAALITDGP-----GGAKR---MYGROQNACTWHDITCN--- 481
QY 352 TSDQAPVAMTDENDLYLSSHNLVNGKEADTAVQGYANPDVSGYLAVWVP 405
DB 482 TGNRSEFPVINSEGWGEFHVNGGSVSIYVQSGEYPAWDFNQIYTNIEIYVHNGQLWQAK 541
QY 406 GASDNQD 412
DB 542 WNTQNE 548

RESULT 10
US-08-812-829-6
Sequence 6, Application US/08812829
Patent No. 6017751
GENERAL INFORMATION:
APPLICANT: von der Osten, Claus
APPLICANT: Bjornvad, Mads E.
APPLICANT: Vind, Jesper
APPLICANT: Rasmussen, Michael Dolberg
TITLE OF INVENTION: PROCESS FOR REMOVAL OR BLEACHING OF SOILING
TITLE OF INVENTION: OR STAINS FROM CELLULOSIC FABRIC
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6017751o No. 6017751disk of No. 6017751th America, Inc.
STREET: 405 Lexington Avenue, Suite 6400
CITY: New York
STATE: New York
COUNTRY: U.S.A.

; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/468,700
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 252
; ATTORNEY/AGENT INFORMATION:
; NAME: Stone, Christopher
; REGISTRATION NUMBER: 35,696
; REFERENCE/DOCKET NUMBER: GC275
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 742-7555
; TELEFAX: (415) 742-7217
; INFORMATION FOR SEQ ID NO: 34:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 483 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-468-700-34

Query Match 5.5%; Score 156.5; DB 1; Length 483;
Best Local Similarity 22.6%; Pred. No. 2.4e-05;
Matches 77; Conservative 40; Mismatches 119; Indels 105; Gaps 14;

QY 7 EFLANDIDNSPVVQAEQLMWLYLLNFGTITANNQANFDSVRVDPADNIDALMNIA 66
DB 194 DYLMTADIDYDHPDVAARIRKW-----GTWYAN--ELQLDGFRLDV----- 233
QY 67 QDYFNAAYGMSDAVSNNKHI--NILEDW-NHADPEYFNKIGNPOLTMDDTIKNSL----- 118
DB 234 -----KHIFSFRLDWNHVR-----EKTGKEMFTVAEYWNQDLAGALEN 272
QY 119 -----NHGLSDATNRGLDAIVHQSLADRENNSTENVIPNYSFVRADNNSQDQION 171
DB 273 YLNKTNENHSVDFVPLHYQFHA-----ASTQGGGYDMRKLNGTVVSKH----- 316
QY 172 AIRDVTGKDYHTFTFDEBQKIDAYIQDNSTVKKYNLYNPASVAILLTNNKDTIPRVY 231
DB 317 PLKSVTFVDNH-----DTQPG-----QSLESTVQTWFK---PLAYAFILITRESGYPOVY 363
QY 232 GDLYTDGGQYMEHQTRYDYDTLTNLLKSRVKYVAGQ-----SMQTMVSVGNNN 279
DB 364 GDMYTKGDSQREIPALKHKTEPIKARKQYAYGAQHDFDHHDIVGWTREGDSSVANS 423
QY 280 ILTSVRYKGA-----MTATDGTDETRTQGIQGVVSN 312
DB 424 LAALITDGPGRKMYGVQRNAGETWHDITGNRSEPVVINS 464

RESULT 13
US-08-468-220-32
; Sequence 32, Application us/08468220
; Patent No. 5824532
; GENERAL INFORMATION:
; APPLICANT: Antrim, Richard L.
; APPLICANT: Barnett, Christopher
; APPLICANT: Mitchinson, Colin
; APPLICANT: Power, Scott D.
; APPLICANT: Regardt, Carol
; APPLICANT: Solheim, Leif P.
; TITLE OF INVENTION: Oxidatively Stable Alpha-Amylase
; NUMBER OF SEQUENCES: 68
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genencor International, Inc.
; STREET: 180 Kimball Way
; CITY: South San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 MB

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/468,220
; FILING DATE: 06-JUN-95
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/194,664
; FILING DATE: 10-FEB-94
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/016,395
; FILING DATE: 11-FEB-93
; ATTORNEY/AGENT INFORMATION:
; NAME: Stone, Christopher L.
; REGISTRATION NUMBER: 35,696
; REFERENCE/DOCKET NUMBER: GC220D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 742-7555
; TELEFAX: (415) 742-7217
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 483 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-468-220-32

Query Match 5.5%; Score 156.5; DB 2; Length 483;
Best Local Similarity 22.6%; Pred. No. 2.4e-05;
Matches 77; Conservative 40; Mismatches 119; Indels 105; Gaps 14;
QY 7 EFLANDIDNSPVVQAEQLMWLYLLNFGTITANNQANFDSVRVDPADNIDALMNIA 66
DB 194 DYLMTADIDYDHPDVAARIRKW-----GTWYAN--ELQLDGFRLDV----- 233
QY 67 QDYFNAAYGMSDAVSNNKHI--NILEDW-NHADPEYFNKIGNPOLTMDDTIKNSL----- 118
DB 234 -----KHIFSFRLDWNHVR-----EKTGKEMFTVAEYWNQDLAGALEN 272
QY 119 -----NHGLSDATNRGLDAIVHQSLADRENNSTENVIPNYSFVRADNNSQDQION 171
DB 273 YLNKTNENHSVDFVPLHYQFHA-----ASTQGGGYDMRKLNGTVVSKH----- 316
QY 172 AIRDVTGKDYHTFTFDEBQKIDAYIQDNSTVKKYNLYNPASVAILLTNNKDTIPRVY 231
DB 317 PLKSVTFVDNH-----DTQPG-----QSLESTVQTWFK---PLAYAFILITRESGYPOVY 363
QY 232 GDLYTDGGQYMEHQTRYDYDTLTNLLKSRVKYVAGQ-----SMQTMVSVGNNN 279
DB 364 GDMYTKGDSQREIPALKHKTEPIKARKQYAYGAQHDFDHHDIVGWTREGDSSVANS 423
QY 280 ILTSVRYKGA-----MTATDGTDETRTQGIQGVVSN 312
DB 424 LAALITDGPGRKMYGVQRNAGETWHDITGNRSEPVVINS 464

RESULT 14
US-08-468-698-32
; Sequence 32, Application US/08468698
; Patent No. 5849549
; GENERAL INFORMATION:
; APPLICANT: Antrim, Richard L.
; APPLICANT: Barnett, Christopher
; APPLICANT: Mitchinson, Colin
; APPLICANT: Power, Scott D.
; APPLICANT: Regardt, Carol
; APPLICANT: Solheim, Leif P.
; TITLE OF INVENTION: Oxidatively Stable Alpha-Amylase
; NUMBER OF SEQUENCES: 68
; CORRESPONDENCE ADDRESS:

;; APPLICANT: GENENCOR INTERNATIONAL, INC.

APPLICANT: Bisg rd-Frantzen, Henrik
APPLICANT: Borchert, Torben Vedel
TITLE OF INVENTION: '-Amylase Mutants
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 5989169o No. 5989169disk of No. 5989169th America, Inc.
STREET: 405 Lexington Avenue, 64th Floor
CITY: New York
STATE: New York
ZIP: 10174-6401
COUNTRY: United States of America
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/600,908A
FILING DATE: 13-FEB-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Green, Reza
REGISTRATION NUMBER: 38,475
REFERENCE/DOCKET NUMBER: 4394,204-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 483 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-600-908A-13

Query Match 5.5%; Score 156.5; DB 2; Length 483;
Best Local Similarity 21.9%; Pred. No. 2.4e-05;
Matches 72; Conservative 46; Mismatches 130; Indels 81; Gaps 13;
QY 7 EFLANDIDNSPVVQAEQLNLYLLNFGTITANNQDQANFDSVRVDAPDNDADLMNTA 66
Db 194 DYLMADVDYDHPDVAEPKKGWIYAN-----ELSLGCFRIDAARHKIK---FSL 241
QY 67 QDYFNAAYGMSDAVSNKHINILEDH--NIAD--PEYFNKIGNPQLTMDTIRNSLNHGLS 123
Db 242 RDWQAV---EQATGKEMPTVAEYQNNAGKLENYLN-----KTSFNQSVF 284
QY 124 DADNRGLDAIVHQSADRENNSSTENVIPNYSFVRAHDNNSQDQIQNAIRDVTGKDIHT 183
Db 285 DVPLEHFNLA-----ASSGGGYDMRKLLNGTVVSKH-----PLKSVTFVDNH- 327
QY 184 FTPEDEQKIDAYIQDQNSVKKYKLYNIPASVAILLTNKTIPRVYVGLYTDGQGYME 243
Db 328 ----DTQPG-----QSLESTVQWFK---PLAYAFILTRREGSGYPQVFGDMYTKGDSQR 375
QY 244 HQPYRYDTLNLKSRVYVAGQ-----SMQTMSVGGNNILTSVRYKGA- 290
Db 376 EIPALKHKIEPILKARKYAGAHQDFDHDIVGWTRREGSDSVANSGLAALITDGPGA 435
QY 291 -----MTATDGTDETRTQIGIGVVSNN 312
Db 436 KRMVYEQNAGETWHDITGNRSEPVVINS 464

RESULT 17
US-08-890-383-3
; Sequence 3, Application US/08890383
; Patent No. 6006026
; GENERAL INFORMATION:
; APPLICANT: Anthony G. Day
; TITLE OF INVENTION: MUTANT ALPHA-AMYLASE HAVING INTRODUCED

TITLE OF INVENTION: THEREIN A DISULFIDE BOND
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genencor International, Inc.
STREET: 925 Page Mill Road
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304-1013
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/890,383
FILING DATE: To Be Assigned
ATTORNEY/AGENT INFORMATION:
NAME: Stone, Christopher
REGISTRATION NUMBER: 35,696
REFERENCE/DOCKET NUMBER: GC377
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 846-7555
TELEFAX: (650) 845-6504
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 483 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-890-383-3

Query Match 5.5%; Score 156.5; DB 3; Length 483;
Best Local Similarity 22.6%; Pred. No. 2.4e-05;
Matches 77; Conservative 40; Mismatches 119; Indels 105; Gaps 14;
QY 7 EFLANDIDNSPVVQAEQLNLYLLNFGTITANNQDQANFDSVRVDAPDNDADLMNTA 66
Db 194 DYLMADVDYDHPDVAEPKKGWIYAN-----GTWYAN--ELQLDGFELDAV----- 233
QY 67 QDYFNAAYGMSDAVSNKHINILEDH--NIADPEYFNKIGNPQLTMDTIRNSLN- 118
Db 234 -----KHKEFSFLRDWVNHVR-----EKTGKEMFTVAEYQNDLGALEN 272
QY 119 -----NHGLSDATNRWGLDAIVHQSADRENNSSTENVIPNYSFVRAHDNNSQDQIQN 171
Db 273 YLKNKTNFNHSFVDPVPLHYQFHA-----ASTGGGYDMRKLLNGTVVSKH----- 316
QY 172 AIRDVTGKDYHTFEDQKIDAYIQDQNSVKKYKLYNIPASVAILLTNKTIPRVYV 231
Db 317 PLKSVTFVDNH-----DTQPG-----QSLESTVQWFK---PLAYAFILTRREGSGYPQV 363
QY 232 GDLYNPGQGYMEHQTRYDTLNLKSRVYVAGQ-----SMQTMSVGGNN 279
Db 364 GDMYTKGDSQREIPALKHKIEPILKARKYAGAHQDFDHDIVGWTRREGSDSVANS 423
QY 280 ILTSVRYKGA-----MTATDGTDETRTQIGIGVVSNN 312
Db 424 LAALITDGPQAKRMVYQNNAGETWHDITGNRSEPVVINS 464

RESULT 18
US-08-683-838A-13
; Sequence 13, Application US/08683838A
; Patent No. 6022724
; GENERAL INFORMATION:
; APPLICANT: Svendsen, Allan
; APPLICANT: Bisg rd-Frantzen, Henrik
; APPLICANT: Borchert, Torben Vedel
; TITLE OF INVENTION: '-Amylase Mutants
; NUMBER OF SEQUENCES: 13

TITLE OF INVENTION: AT RESIDUES CORRESPONDING TO A210, H405 AND/OR
 TITLE OF INVENTION: T412 IN BACILLUS LICHENIFORMIS
 NUMBER OF SEQUENCES: 12
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Genencor International, Inc.
 STREET: 925 Page Mill Road
 CITY: Palo Alto
 STATE: CA
 COUNTRY: USA
 ZIP: 94304-1013
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FastSeq for Windows Version 2.0
 CURRENT APPLICATION DATA: US/08/914,679A
 FILING DATE: To Be Assigned
 ATTORNEY/AGENT INFORMATION:
 NAME: Stone, Christopher
 REGISTRATION NUMBER: 35,696
 REFERENCE/DOCKET NUMBER: GC387
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (650) 846-7535
 TELEFAX: (650) 845-6504
 INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 483 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: unknown
 MOLECULE TYPE: protein
 PS-08-914-679A-3

[illegible]

```

RESULT 20
US-09-182-859-2
; Sequence 2, Application US/09182859
; Patent No. 6143708
; GENERAL INFORMATION:
; APPLICANT: Svendsen, Allan
; APPLICANT: Borchert, Torben
; APPLICANT: Bisgard-Frantzen, Henrik
; TITLE OF INVENTION: Alpha-Amylase Mutants

```

FILE REFERENCE: 4796.204-US
; CURRENT APPLICATION NUMBER: US/09/182.859
; CURRENT FILING DATE: 1998-10-29
; EARLIER APPLICATION NUMBER: 0515/96
; EARLIER FILING DATE: 1996-04-30
; EARLIER APPLICATION NUMBER: 0712/96
; EARLIER FILING DATE: 1996-06-28
; EARLIER APPLICATION NUMBER: 0775/96
; EARLIER FILING DATE: 1996-07-11
; EARLIER APPLICATION NUMBER: 1263/96
; EARLIER FILING DATE: 1996-11-08
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 483
; TYPE: PRT
; ORGANISM: Bacillus licheniformis
US-09-182-859-2

Query Match 5.5%; Score 156.5; DB 4; Length 483;
Best Local Similarity 22.6%; Pred. No. 2.4e-05;
Matches 77; Conservative 40; Mismatches 119; Indels 105; Gaps 14;
QY 7 EFLANDIDNSNPVQAEOQLWLYLLNFGTITANNQANFDSVRVDAPDNIDADLMNIA 66
Db 194 DYLMAADIDYDHPDVAAEIKRW-----GTWYAN--ELQDGFRLDAV----- 233
QY 67 QDYFNAAYGMSDAVSNKHI--NILEDW-NHADPEYFNKIGNPQLTMDDTIKNSL----- 118
Db 234 -----KHIFSFRLDWNHVR-----EKTGKEMFTVAEYQNDLGALEN 272
QY 119 -----NHGLSDATNRWGLDAIVHQSLADRENNSTENVIPNYSFVRAHDNNSQDQION 171
Db 273 YLKNKTNFNHSFVDFPLHYQFHA-----ASTQGGGYDMRKLINGTVVSKH----- 316
QY 172 AIRDVTGKDYHTFFEDQKGDIDAYIQDNSTVKYKYNLYNPASVAILLTNKTIPRVYY 231
Db 317 PLKSVTFVDNH-----DTQPG-----QSLESTVQWTFK---PLAYAFILTRSGYPQVY 363
QY 232 GDLYTDGGQYMEHQTRYDITLNLKSRVKYVAGQ-----SMQTMVSGGNN 279
Db 364 GDMYGTGDSQREIPALKHKIEPILKARKQYAGAHYDFDHDHIVGWTREGDSSVANS 423
QY 280 ILTSVRYKGA-----MTATDTGDETRTQIGVGVVSN 312
Db 424 LAALITDGGGAKRMVYGRQNAGETWHDITGNRSEPVVINS 464

RESULT 21
US-09-170-670-4
; Sequence 4, Application US/09170670
; Patent No. 6187576
; GENERAL INFORMATION:
; APPLICANT: Svendsen, Allan
; APPLICANT: Borchert, Torben
; APPLICANT: Bisgaard-Frantzen Henrik
; TITLE OF INVENTION: Alpha-Amylase Mutants
; FILE REFERENCE: 5276.200-US
; CURRENT APPLICATION NUMBER: US/09/170,670
; CURRENT FILING DATE: 1998-10-13
; EARLIER APPLICATION NUMBER: 1172/97
; EARLIER FILING DATE: 1997-10-13
; EARLIER APPLICATION NUMBER: 60/063,306
; EARLIER FILING DATE: 1997-10-28
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 483
; TYPE: PRT
; ORGANISM: B. licheniformis
US-09-170-670-4

Query Match 5.5%; Score 156.5; DB 4; Length 483;
Best Local Similarity 22.6%; Pred. No. 2.4e-05;
Matches 77; Conservative 40; Mismatches 119; Indels 105; Gaps 14;
QY 7 EFLANDIDNSNPVQAEOQLWLYLLNFGTITANNQANFDSVRVDAPDNIDADLMNIA 66
Db 194 DYLMAADIDYDHPDVAAEIKRW-----GTWYAN--ELQDGFRLDAV----- 233
QY 67 QDYFNAAYGMSDAVSNKHI--NILEDW-NHADPEYFNKIGNPQLTMDDTIKNSL----- 118
Db 234 -----KHIFSFRLDWNHVR-----EKTGKEMFTVAEYQNDLGALEN 272
QY 119 -----NHGLSDATNRWGLDAIVHQSLADRENNSTENVIPNYSFVRAHDNNSQDQION 171
Db 273 YLKNKTNFNHSFVDFPLHYQFHA-----ASTQGGGYDMRKLINGTVVSKH----- 316
QY 172 AIRDVTGKDYHTFFEDQKGDIDAYIQDNSTVKYKYNLYNPASVAILLTNKTIPRVYY 231
Db 317 PLKSVTFVDNH-----DTQPG-----QSLESTVQWTFK---PLAYAFILTRSGYPQVY 363
QY 232 GDLYTDGGQYMEHQTRYDITLNLKSRVKYVAGQ-----SMQTMVSGGNN 279
Db 364 GDMYGTGDSQREIPALKHKIEPILKARKQYAGAHYDFDHDHIVGWTREGDSSVANS 423
QY 280 ILTSVRYKGA-----MTATDTGDETRTQIGVGVVSN 312
Db 424 LAALITDGGGAKRMVYGRQNAGETWHDITGNRSEPVVINS 464

RESULT 22
US-09-193-068-4
; Sequence 4, Application US/09193068
; Patent No. 6197565
; GENERAL INFORMATION:
; APPLICANT: Svendsen, Allan
; APPLICANT: Kjruliff, S ren
; APPLICANT: Bisgaard-Frantzen, Henrik
; APPLICANT: Andersen, Carsten
; TITLE OF INVENTION: -Amylase Variants
; FILE REFERENCE: 5709.000-US
; CURRENT APPLICATION NUMBER: US/09/193,068
; CURRENT FILING DATE: 1998-11-16
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 483
; TYPE: PRT
; ORGANISM: B. licheniformis
US-09-193-068-4

Query Match 5.5%; Score 156.5; DB 4; Length 483;
Best Local Similarity 22.6%; Pred. No. 2.4e-05;
Matches 77; Conservative 40; Mismatches 119; Indels 105; Gaps 14;
QY 7 EFLANDIDNSNPVQAEOQLWLYLLNFGTITANNQANFDSVRVDAPDNIDADLMNIA 66
Db 194 DYLMAADIDYDHPDVAAEIKRW-----GTWYAN--ELQDGFRLDAV----- 233
QY 67 QDYFNAAYGMSDAVSNKHI--NILEDW-NHADPEYFNKIGNPQLTMDDTIKNSL----- 118
Db 234 -----KHIFSFRLDWNHVR-----EKTGKEMFTVAEYQNDLGALEN 272
QY 119 -----NHGLSDATNRWGLDAIVHQSLADRENNSTENVIPNYSFVRAHDNNSQDQION 171
Db 273 YLKNKTNFNHSFVDFPLHYQFHA-----ASTQGGGYDMRKLINGTVVSKH----- 316
QY 172 AIRDVTGKDYHTFFEDQKGDIDAYIQDNSTVKYKYNLYNPASVAILLTNKTIPRVYY 231
Db 317 PLKSVTFVDNH-----DTQPG-----QSLESTVQWTFK---PLAYAFILTRSGYPQVY 363
QY 232 GDLYTDGGQYMEHQTRYDITLNLKSRVKYVAGQ-----SMQTMVSGGNN 279

```

Db 364 GDMYGTGDSQREIPALKHKIEPILKARKQAYGAQHDFDHHDIVGWTREGDSSVANS 423
QY 280 ILTSVRVYKGA-----MTADTGTDETRTQGGIGVVVSN 312
Db 424 LAALITDGGGAKRMVYGRONAGETWHDITGNRSEPVVINS 464

RESULT 23
US-09-183-412-4
; Sequence 4, Application US/09183412
; Patent No. 6204232
; GENERAL INFORMATION:
; APPLICANT: Borchert, Torben V.
; APPLICANT: Svendsen, Allan
; APPLICANT: Andersen, Carsten
; APPLICANT: Nielsen, Bjarne
; APPLICANT: Nissen, Torben L.
; APPLICANT: Kiaeruliff, Soren
; TITLE OF INVENTION: Alpha-Amulase Mutants
; FILE REFERENCE: 5368.200-US
; CURRENT APPLICATION NUMBER: US/09/183,412
; EARLIER FILING DATE: 1998-10-30
; EARLIER APPLICATION NUMBER: 60/064,662
; EARLIER FILING DATE: 1997-11-06
; EARLIER APPLICATION NUMBER: 60/093,234
; EARLIER FILING DATE: 1998-07-17
; EARLIER APPLICATION NUMBER: 1240/97
; EARLIER FILING DATE: 1997-10-30
; EARLIER APPLICATION NUMBER: PA 1998 00936
; EARLIER FILING DATE: 1998-07-14
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 483
; TYPE: PRT
; ORGANISM: Bacillus licheniformis
US-09-183-412-4

Query Match 5.5%; Score 156.5; DB 4; Length 483;
Best Local Similarity 22.6%; Pred. No. 2.4e-05;
Matches 77; Conservative 40; Mismatches 119; Indels 105; Gaps 14;

QY 7 EFLANDIDNSNPVQAQQLNMLYLLNFGTITANNQANFDSVRVDPADNIDLMLNIA 66
Db 194 DYLMYADIDYDHPDVAAEIKRW-----GTWYAN--ELQDGFRLDAV----- 233
QY 67 QDYFNAAYGMSDAVSNKHI--NILEDW-NHADPEYFNKIGNPQLTMDDTIKNSL----- 118
Db 234 -----KHKFSFLRDWVNHVR-----EKTGKEMFTVAEYQNDLGALEN 272
QY 119 -----NHGLSDATNRWGLDAIVHQSGLADRENNSSTENVVPIYSFVRAHDNNSQDQIQN 171
Db 273 YLNKTFNHSVDFVPLHYQFHA-----ASTGGGCDMRKLLNGTVVSKH----- 316
QY 172 AYRDVTGKDYHTFTFEDEQKIDAYIQDQNSTVKKYNLYNPASVAILLTNKTIPRVY 231
Db 317 PLKSVTFVDNH-----DTQPG-----QSESTVQTWFK---PLAYAFILTRSGYPOVY 363
QY 232 GDLYTDGGQYMEHQTRYDPTLTNLLKSRVKYVAGGQ-----SMQTSVSGGN 279
Db 364 GDMYGTGDSQREIPALKHKIEPILKARKQAYGAQHDFDHHDIVGWTREGDSSVANS 423
QY 280 ILTSVRVYKGA-----MTADTGTDETRTQGGIGVVVSN 312
Db 424 LAALITDGGGAKRMVYGRONAGETWHDITGNRSEPVVINS 464

RESULT 24
US-09-985-659-35
; Sequence 35, Application US/09985659
; Patent No. 6211134
```

```

; GENERAL INFORMATION:
; APPLICANT: Caldwell, Robert M
; APPLICANT: Mitchinson, Colin
; APPLICANT: Ropp, Traci M
; TITLE OF INVENTION: Mutant Alpha-Amylase
; FILE REFERENCE: A-69396/DJB/DAV/JJD
; CURRENT APPLICATION NUMBER: US/08/985,659
; CURRENT FILING DATE: 1997-12-09
; PRIOR APPLICATION NUMBER: 08/645,971
; PRIOR FILING DATE: 1996-05-14
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 35
; LENGTH: 483
; TYPE: PRT
; ORGANISM: Bacillus licheniformis
US-08-985-659-35

Query Match 5.5%; Score 156.5; DB 4; Length 483;
Best Local Similarity 22.6%; Pred. No. 2.4e-05;
Matches 77; Conservative 40; Mismatches 119; Indels 105; Gaps 14;

QY 7 EFLANDIDNSNPVQAQQLNMLYLLNFGTITANNQANFDSVRVDPADNIDLMLNIA 66
Db 194 DYLMYADIDYDHPDVAAEIKRW-----GTWYAN--ELQDGFRLDAV----- 233
QY 67 QDYFNAAYGMSDAVSNKHI--NILEDW-NHADPEYFNKIGNPQLTMDDTIKNSL----- 118
Db 234 -----KHKFSFLRDWVNHVR-----EKTGKEMFTVAEYQNDLGALEN 272
QY 119 -----NHGLSDATNRWGLDAIVHQSGLADRENNSSTENVVPIYSFVRAHDNNSQDQIQN 171
Db 273 YLNKTFNHSVDFVPLHYQFHA-----ASTGGGCDMRKLLNGTVVSKH----- 316
QY 172 AYRDVTGKDYHTFTFEDEQKIDAYIQDQNSTVKKYNLYNPASVAILLTNKTIPRVY 231
Db 317 PLKSVTFVDNH-----DTQPG-----QSESTVQTWFK---PLAYAFILTRSGYPOVY 363
QY 232 GDLYTDGGQYMEHQTRYDPTLTNLLKSRVKYVAGGQ-----SMQTSVSGGN 279
Db 364 GDMYGTGDSQREIPALKHKIEPILKARKQAYGAQHDFDHHDIVGWTREGDSSVANS 423
QY 280 ILTSVRVYKGA-----MTADTGTDETRTQGGIGVVVSN 312
Db 424 LAALITDGGGAKRMVYGRONAGETWHDITGNRSEPVVINS 464

RESULT 25
US-09-264-097-2
; Sequence 2, Application US/09264097
; Patent No. 6287826
; GENERAL INFORMATION:
; APPLICANT: No. 6287826man, Barrie Edmund
; APPLICANT: Hendriksen, Hanne Vang
; TITLE OF INVENTION: Enzymatic Preparation of Glucose Syrup
; FILE REFERENCE: 5278.200-US
; CURRENT APPLICATION NUMBER: US/09/264,097
; CURRENT FILING DATE: 1999-03-08
; EARLIER APPLICATION NUMBER: PA 0321/98
; EARLIER FILING DATE: 1998-03-09
; EARLIER APPLICATION NUMBER: 60/079,209
; EARLIER FILING DATE: 1998-03-24
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 483
; TYPE: PRT
; ORGANISM: Bacillus licheniformis
US-09-264-097-2
```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 12, 2002, 07:56:08 ; Search time 20.97 Seconds

(without alignments)
2488.150 Million cell updates/sec

Title: US-09-995-749A-2_COPY_972_1514

Perfect score: 2852

Sequence: 1 QTVRGQBEFLANDINSNPV.....AIQALHKAGLQVMDWVPDQ 543

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

1: PIR1:*

2: PIR2:*

3: PIR3:*

4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1464.5	51.3	1375	2 JT0345	dextranucrase (EC
2	1464.5	51.3	1508	2 T31098	probable dextran
3	1413	49.5	1431	2 A45866	dextranucrase (EC
4	1407.5	49.4	1449	2 T30857	glucosyltransferas
5	1407.5	49.4	1449	2 T30552	glucosyltransferas
6	1406	49.3	1592	2 A38175	glucosyltransferas
7	1397	49.0	1475	2 B33135	glfB protein precu
8	1377	48.3	1290	2 JC5473	dextranucrase (EC
9	1288.5	45.2	1577	2 T30858	glucosyltransferas
10	1231.5	43.2	1518	2 A44811	glucosyltransferas
11	1219.5	42.8	1365	2 A41483	glucosyltransferas
12	1179.5	41.4	1599	2 S22737	glucosyltransferas
13	160.5	5.6	1385	2 D89824	hypothetical prote
14	155.5	5.5	512	1 A488L	alpha-amylase (EC
15	152.5	5.3	1315	2 S578679	fibrinogen-binding
16	151.5	5.3	601	2 S57962	cspC protein - Clo
17	150	5.3	1377	2 D90538	hypothetical prote
18	149.5	5.2	514	1 ALBSN	alpha-amylase (EC
19	149.5	5.2	2348	2 AD1841	hypothetical prote
20	147	5.2	4152	2 T31102	filamentous hemagg
21	145.5	5.1	1148	2 S72635	exo-poly-alpha-gal
22	144.5	5.1	518	1 A27705	alpha-amylase (EC
23	144	5.0	6713	2 B89921	hypothetical prote
24	143	5.0	584	2 G71676	hypothetical prote
25	141.5	5.0	718	1 ALBSCC	cyclomaltodextrin
26	141.5	5.0	786	2 T18469	hypothetical prote
27	141	4.9	495	1 A45738	alpha-amylase (EC
28	139.5	4.9	713	2 A58800	cyclomaltodextrin
29	139.5	4.9	1487	2 AG2560	hypothetical prote

30	139	4.9	495	2 B90962	cytoplasmic alpha-
31	139	4.9	495	2 B85810	cytoplasmic alpha-
32	139	4.9	634	2 F97172	flagellar hook-ass
33	139	4.9	3890	2 C89921	hypothetical prote
34	137.5	4.8	1588	2 A86036	probable adhesin 2
35	137.5	4.8	1588	2 H91188	probable adhesin 3
36	137	4.8	566	1 HVBSU	bacillolysin (EC 3
37	137	4.8	1902	1 B44858	lactocarpin (EC 3.4
38	136	4.8	1026	1 TLBP74	tail fiber protein
39	136	4.8	1578	2 AD1512	peptidoglycan boun
40	134.5	4.7	461	2 S42084	basic protein - Cl
41	134.5	4.7	1335	2 T30211	autolysin E - Stap
42	134	4.7	718	1 ALBSG6	cyclomaltodextrin
43	134	4.7	4936	2 AH2515	hypothetical prote
44	133.5	4.7	953	2 C89824	hypothetical prote
45	133.5	4.7	1021	2 I40805	collagenase - C1os
46	133.5	4.7	1072	2 A86827	hypothetical prote
47	133.5	4.7	1819	2 D97033	uncharacterized pr
48	133	4.7	1313	2 G82887	hypothetical prote
49	133	4.7	2500	2 G71609	hypothetical prote
50	132.5	4.6	484	2 F98026	alpha-amylase (EC
51	132.5	4.6	3029	2 S76109	hypothetical prote
52	131.5	4.6	1993	2 AF1450	probable peptidogl
53	131	4.6	498	2 G97279	protein containing
54	131	4.6	548	1 ALBSF	alpha-amylase (EC
55	131	4.6	694	2 F97279	TPR-repeat-contain
56	130.5	4.6	439	2 G97159	contains cell adhe
57	130.5	4.6	484	2 G95160	alpha-amylase [imp
58	130.5	4.6	583	2 S57721	cspB protein - Clo
59	130.5	4.6	793	2 S73662	probable lipoprote
60	130.5	4.6	2271	2 F90073	hypothetical prote
61	130.5	4.6	2364	2 I40884	cytotoxin L - C1os
62	129.5	4.5	1666	2 T28680	fibrinogen-binding
63	129.5	4.5	1260	2 H89984	hypothetical prote
64	129	4.5	581	2 S62505	probable alpha-am
65	129	4.5	826	2 E96720	probable alpha-amy
66	129	4.5	852	2 A28459	cell surface glyco
67	129	4.5	1349	2 A11476	cell surface prote
68	128.5	4.5	1475	2 A44765	alpha-amylase (EC
69	128	4.5	549	1 A54541	alpha-amylase (EC
70	128	4.5	549	1 A24549	alpha-amylase (EC
71	128	4.5	549	1 A24436	alpha-amylase (EC
72	128	4.5	731	2 B70770	probable glgB prot
73	128	4.5	1296	2 C82521	hemolysin-type cal
74	128	4.5	1645	2 F96907	phage-related prot
75	128	4.5	1902	2 B45764	lactocarpin (EC 3.4
76	128	4.5	2334	2 S32920	cell wall-associat
77	127.5	4.5	712	1 ALBSG3	cyclomaltodextrin
78	127.5	4.5	1536	2 A43855	high-molecular-wel
79	127.5	4.5	1592	2 AC1153	adhesin homolog lm
80	127	4.5	600	2 G83081	conserved hypothet
81	127	4.5	1300	2 S07575	outer membrane pro
82	127	4.5	1403	2 S77624	mannuronan C-5-epi
83	127	4.5	2020	2 C48399	ABC-type transport
84	127	4.5	2366	2 S10317	toxin B - Clostrid
85	126.5	4.4	611	2 JC7177	endoglucanase V [E
86	126.5	4.4	622	2 AF0169	probable exported
87	126.5	4.4	713	2 S09196	cyclomaltodextrin
88	126	4.4	465	2 H97165	flagellar hook-len
89	126	4.4	980	2 H90681	probable flagellin
90	126	4.4	980	2 D85532	probable structura
91	126	4.4	1092	2 T30214	fibrinogen-binding
92	126	4.4	1883	2 G82875	hypothetical prote
93	126	4.4	2555	2 AC0304	probable hemolysin
94	125.5	4.4	507	2 S52348	hypothetical prote
95	125.5	4.4	818	2 F82173	collagenase VC1650
96	125.5	4.4	1107	2 AC0976	probable autotrans
97	125	4.4	429	1 JE0406	alpha-amylase (EC
98	125	4.4	435	2 AD3038	alpha-amylase amya
99	125	4.4	506	2 G98247	cytoplasmic alpha-
100	125	4.4	584	2 C48658	flagellin - Escher

Db 944 QIAPQYRSTDTSTFLDSIIQNGYAFTRDYDLGYPPTKYGTVDQLRDAIKALHANGIOAI 1003
QY 537 ADWVPDQ 543
| | | | |
Db 1004 ADWVPDQ 1010

RESULT 3
A:Accession: A45866
C:Species: Streptococcus mutans
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 07-May-1999
C:Accession: A45866
R:Simpson, C.L.: Giffard, P.M.: Jacques, N.A.
Infected. Immun. 63, 609-621, 1995

A:Title: Streptococcus mutans gtfd gene encoding the glucosyltransferase (EC 2.4.1.5) precursor - Streptococcus mutans
A:Reference number: 220909; MUID:95122197
A:Accession: T30857

A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1449 <SIM>
A:Cross-references: EMBL:L35495; NID:g662378; PID:g662379; PIDN:AA041412.1
C:Genetics:
A:Gene: gtfI

Query Match 49.5%; Score 1413; DB 2; Length 1431;
Best Local Similarity 50.3%; Pred. No. 6.7e-78;
Matches 282; Conservative 96; Mismatches 149; Indels 34; Gaps 8;
QY 5 GQEFLLANDINSNPVQAEQLNWLIIYLNFTITANNDOANFDSVRVDADPNIDADLMN 64
Db 417 GYDFLLANDINSNPVQAEQLNWLIIYLNFTITANNDOANFDSVRVDADPNIDADLMN 476

QY 65 IAQDYFNAAYGMD-SDAVSNKHINLEADNADPEYFNKIGNPQLTMDTITKNSLNHGLS 123
Db 477 IASDYLUKHYGVDSKSEKNAHINLSLEADNADPEYFNKIGNPQLTMDTITKNSLNHGLS 536

QY 124 -----DATN----RWGLDAIVHQSADRENNSSTENVIPYSPVRAHDNNSQDTONAIR 174
Db 537 RPLEKASNKNEIRSGLEPVITNSLNRSAGKNSERMANIYFIRAHDSVQTVIAKIK 596

QY 175 -DVGK-DYHTFFDEQKIDAYIQDNSTVKKNLYNPASYAILLTNKTIPRYVYGDLYTD 232
Db 597 AQINPKTDGLTDLDELKQAFIYNEDMRQAKKYYQSNIPYALMLNKNKDSITRLYYG 656

QY 233 DLYTDGQYMEHQTRYDITLTLNLSKRYVAGGQSMQTSVGGNNN-----ILTSV 284
Db 657 DMVSDGQYMAKSPYDAILTLKARIKAGGQDMKITYEGDKSHMDWYTGVLTSV 716

QY 285 RYKGAWTATDGTDETRTQIGGVVSYNTPNLKLGVNDKVLVHMGAAHKNQYRAAVLPT 344
Db 717 RYGTGANEATDQSEATKTQGMVITSNP SLKQNDKVIVNAGAAHKNQYRPLILT 776

QY 345 TDGVINTSDGAP--VAMTDENGDLVLSHNLVNGKEADTAVQGYANPDVSGYLAVW 402
Db 777 KGLTSTYSDAAKSLRYKNTDKGELVFDASD-----IQGYLNPQVSGYLAVW 824

QY 403 VPVGASDNODARTAPSTKNSGNSAYRTNAAPDSNVIFEAFSNFYVTPTKESERANVRIA 462
Db 825 VPVGASDNQDVRVAASNAKATQGVYESSALSQLIYEGFSNFQDFVTKDSYTNKXIA 884

QY 463 QNADFFASIGFTSFEMAPQYNSKDRFTFLDSTIDNGYAFTRDYDLGSEPNKYGTDEDLR 522
Db 885 QNVLPKSMGVTSFEMAPQYVSSDGSFLDSIIQNGYAFTRDYDLGSEPNKYGTDEDLR 944

QY 523 NAIQALHKAGLQVWADWVPDQ 543
| | | | |
Db 945 NAVKALHKSGIQVIADWVPDQ 965

RESULT 4
T30857
glucosyltransferase - Streptococcus salivarius
C:Species: Streptococcus salivarius
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
C:Accession: T30857
R:Simpson, C.L.: Giffard, P.M.: Jacques, N.A.
Infected. Immun. 63, 609-621, 1995

A:Title: Streptococcus salivarius ATCC 25975 possesses at least two genes coding for
A:Reference number: 220909; MUID:95122197
A:Accession: T30857

A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1449 <SIM>
A:Cross-references: EMBL:L35495; NID:g662378; PID:g662379; PIDN:AA041412.1
C:Genetics:
A:Gene: gtfI

Query Match 49.4%; Score 1407.5; DB 2; Length 1449;
Best Local Similarity 51.5%; Pred. No. 1.5e-77;
Matches 286; Conservative 81; Mismatches 155; Indels 33; Gaps 9;
QY 5 GQEFLLANDINSNPVQAEQLNWLIIYLNFTITANNDOANFDSVRVDADPNIDADLMN 64
Db 458 GPEFLANDINSNPVQAEQLNWLIIYLNFTITANNDOANFDSVRVDADPNIDADLMN 517

QY 65 IAQDYFNAAYGMD-SDAVSNKHINLEADNADPEYFNKIGNPQLTMDTITKNSLNHGLS 122
Db 518 IASDYLUKHYGVDSKSEKNAHINLSLEADNADPEYFNKIGNPQLTMDTITKNSLNHGLS 577

QY 123 SDATNRGLDAIVHQSADRENNSSTENVIPYSPVRAHDNNSQDTONAIRD-----VT 177
Db 578 RPIGRSGVEPLISLNDRESGKNSKRMANYAFVRAHDSVQSIIGQIKNEINPOST 637

QY 178 GRDYTFTEDEQKIDAYIQDNSTVKKNLYNPASYAILLTNKTIPRYVYGDLYTD 237
Db 638 G---NFTLDEMKAKEIYNKDMRSANKQYTOYNIPSAVALMLTKDTPRVYVYGDLYTD 694

QY 238 GQYMEHQTRYDITLTLNLSKRYVAGGQSMQTSVGGNNN-----ILTSVRYKGA 290
Db 695 DGQYMAQKSPYDAILTLKGRIRYAGGQDMKITYEGNTNGWDAAGVLTSVRYGTGA 754

QY 291 MTATDTGTDDETRTQIGGVVSYNTPNLKLGVNDKVLVHMGAAHKNQYRAAVLTDTGVIN 350
Db 755 NSASDTGTAETRNQGHAVIVSNOPALRLTSN--LTINGGAHKNQYRAVPLLTNDGAT 812

QY 351 YTSQDQAP--VAMTDENGDLVLSHNLVNGKEADTAVQGYANPDVSGYLAVWVPVGS 408
Db 813 YLNDSDANGIVRYTDCNGNLTPSANE-----IRGIRNPDQVGYLAVWVPVGS 860

QY 409 DNQDARTAPSTKNSGNSAYRTNAAPDSNVIFEAFSNFYVTPTKESERANVRIAQADFF 468
Db 861 ENQDVAPSPKNSGLVYESNAALDSQVIYEGFSNFQDFVONPQYTNKKIAENANLF 920

QY 469 ASLGFTSFEMAPQYNSKDRFTFLDSTIDNGYAFTRDYDLGSEPNKYGTDEDLR 528
Db 921 KSWGITSFEAFQYVSSDGSFLDSVIQNGYAFTRDYDLGSEPNKYGTDEDLR 980

QY 529 HKAGLQVWADWVPDQ 543
| | | | |
Db 981 HAVGISAIADWVPDQ 995

RESULT 5
T30552
glucosyltransferase N - Streptococcus salivarius (fragment)

A:Gene: gtfm

Query Match 45.2%; Score 1288.5; DB 2; Length 1577;
Best Local Similarity 48.3%; Pred. No. 2.9e-70;
Matches 265; Conservative 94; Mismatches 163; Indels 27; Gaps 10;

Qy 5 GQFELLANDINSNPVQAEQLNWLYLLNFETITANNQDQANFDSVRVDA PDNIDADLMN 64
Dy 423 GFDELLANDVLSNPVQAEQLNQIHYLMNWSIVMGDKDANFDGIRVDAVDNVDADMLQ 482
Qy 65 IAQDYFNAAYGMD-SDAVSNKHIIINLEDDNHADPEYFNKIGNPOLTMDDTIKNSLNHGLS 123
Dy 483 LYTNIFRYRYGVNKSANLNAHISVLEAWSLNDHNDKTDGAALAMENKQRLALLSLA 542
Qy 124 -----DATNRWGLDAILYHQSLADRENNSTENV 151
Dy 543 KPIKERTPAVSPLYNNIETTORDEKTDWINKQSKAYNEDGTVKQSTIGKYNEKYGD-A 601
Qy 152 IPNYSFVRAHDNNSQDQONAI-RDVTGK-DYHFTFDEQKGDAYIQDQNSVVKYNL 209
Dy 602 SGNVYFIRAHNNVQDIIAEIIRKBNPKSDGFTTDAEMKQAFEIYNKMLSDSKKYYL 661
Qy 210 YNIPASAYAILLTNKTIPRVYVGLYTDGGQYMEHOTRYDITLNLKSRKYVYAGGSM 269
Dy 662 NNIPASAYAVMLQNMETIRVYVGLYTDGGHYMETKSPYDITVNLAKSRKYVSGGQ 721
Qy 270 QT-----MSVGGNNILTSVRYGKAMTATDT-GTDETRTQGIQVVSNTPN 315
Dy 722 RSYMLPTDGMDSVVELYRINEVYTSVRYGKIDINTANDTEGSKYRTSGOVLVANNPK 781
Qy 316 LKLYNDKVLVLMHGAHKNQYRAAVLATTTDGVINYSQDQAPVA---MTDENCOLYLS 371
Dy 782 LMLDSAKLVNEMGKIRANQKTRALIVGTADGIRKFTSDADAIAAGYVKETDSGLVLTG 841
Qy 372 SHNVVNGKEEADTAVQGYANDVSGYLAVVVPVPGASDNQDARTAPSTE-KNSGNSAYRT 430
Dy 842 AND-----IKGYETDMSGFVAVVVPVPGASDNQDILRVAPSTEAKKEGELTKA 889
Qy 431 NNAFOSNVIFRAFSNFVTP--TKESERANVRVIAQNADEFFASLGFTSEMAPQYNSSKDR 488
Dy 890 TEAYDSQILYEGFSNFQIPDGSDFSVTNRKIAENVDLKFSWGVTSFEMAPQFYSADG 949
Qy 489 TFLDSTIDNGYAFTRDYLGMSEPNKYGTDEDLRNAIQALHKAQVQVADVPDQ 543
Dy 950 TFLDSVIQNGYAFADRIDLAMSNNKYSGKEDRLKALHKAQVQVADVPDQ 1004

RESULT 11
A41483
glucosyltransferase (EC 2.4.1.-) gtfS precursor - Streptococcus sobrinus
C:Species: Streptococcus sobrinus
C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 15-Oct-1999
C:Accession: A41483
R:Gilmore, K.S.; Russell, R.R.B.; Ferretti, J.J.
A:Title: Analysis of the Streptococcus downei gtfS gene, which specifies a glucosyltr
A:Reference number: A41483; MUID:90316665
A:Accession: A41483
A:Molecule type: DNA
A:Residues: 1-1365 <GIL>
A:Cross-references: GB:M30943; NID:g153652; PIDN:AAA26898.1; PID:g153653
C:Genetics:
A:Gene: gtfS
C:Superfamily: cpl repeat homology
C:Keywords: glycosyltransferase; hexosyltransferase

Query Match 42.8%; Score 1219.5; DB 2; Length 1365;
Best Local Similarity 46.8%; Pred. No. 3.8e-66;
Matches 259; Conservative 82; Mismatches 181; Indels 31; Gaps 9;

Qy 3 YRQGEFLANDINSNPVQAEQLNWLYLLNFETITANNQDQANFDSVRVDA PDNIDADL 62
Dy 387 YAGYELLANDVDSNPVQAEQLNHLHLMWGDIVMGDKDANFDGIRVDAVDNADL 446
Qy 63 MNTAQDYFNAAYGMD-SDAVSNKHIIINLEDDNHADPEYFNKIGNPOLTMDDTIKNSLNH 121
Dy 447 LQIQDYRYKAKYGTIDQNEKNAIDHLSILEAWSGNDNDYVYKQDNQNFSLSDNDQDSGLKA 506

A:Gene: gtfm

Query Match 45.2%; Score 1288.5; DB 2; Length 1577;
Best Local Similarity 48.3%; Pred. No. 2.9e-70;
Matches 265; Conservative 94; Mismatches 163; Indels 27; Gaps 10;

Qy 5 GQFELLANDINSNPVQAEQLNWLYLLNFETITANNQDQANFDSVRVDA PDNIDADLMN 64
Dy 423 GFDELLANDVLSNPVQAEQLNQIHYLMNWSIVMGDKDANFDGIRVDAVDNVDADMLQ 482
Qy 65 IAQDYFNAAYGMD-SDAVSNKHIIINLEDDNHADPEYFNKIGNPOLTMDDTIKNSLNHGLS 123
Dy 483 LYTNIFRYRYGVNKSANLNAHISVLEAWSLNDHNDKTDGAALAMENKQRLALLSLA 542
Qy 124 -----DATNRWGLDAILYHQSLADRENNSTENV 151
Dy 543 KPIKERTPAVSPLYNNIETTORDEKTDWINKQSKAYNEDGTVKQSTIGKYNEKYGD-A 601
Qy 152 IPNYSFVRAHDNNSQDQONAI-RDVTGK-DYHFTFDEQKGDAYIQDQNSVVKYNL 209
Dy 602 SGNVYFIRAHNNVQDIIAEIIRKBNPKSDGFTTDAEMKQAFEIYNKMLSDSKKYYL 661
Qy 210 YNIPASAYAILLTNKTIPRVYVGLYTDGGQYMEHOTRYDITLNLKSRKYVYAGGSM 269
Dy 662 NNIPASAYAVMLQNMETIRVYVGLYTDGGHYMETKSPYDITVNLAKSRKYVSGGQ 721
Qy 270 QT-----MSVGGNNILTSVRYGKAMTATDT-GTDETRTQGIQVVSNTPN 315
Dy 722 RSYMLPTDGMDSVVELYRINEVYTSVRYGKIDINTANDTEGSKYRTSGOVLVANNPK 781
Qy 316 LKLYNDKVLVLMHGAHKNQYRAAVLATTTDGVINYSQDQAPVA---MTDENCOLYLS 371
Dy 782 LMLDSAKLVNEMGKIRANQKTRALIVGTADGIRKFTSDADAIAAGYVKETDSGLVLTG 841
Qy 372 SHNVVNGKEEADTAVQGYANDVSGYLAVVVPVPGASDNQDARTAPSTE-KNSGNSAYRT 430
Dy 842 AND-----IKGYETDMSGFVAVVVPVPGASDNQDILRVAPSTEAKKEGELTKA 889
Qy 431 NNAFOSNVIFRAFSNFVTP--TKESERANVRVIAQNADEFFASLGFTSEMAPQYNSSKDR 488
Dy 890 TEAYDSQILYEGFSNFQIPDGSDFSVTNRKIAENVDLKFSWGVTSFEMAPQFYSADG 949
Qy 489 TFLDSTIDNGYAFTRDYLGMSEPNKYGTDEDLRNAIQALHKAQVQVADVPDQ 543
Dy 950 TFLDSVIQNGYAFADRIDLAMSNNKYSGKEDRLKALHKAQVQVADVPDQ 1004

RESULT 10
A44811
glucosyltransferase (EC 2.4.1.-) I - Streptococcus salivarius
C:Species: Streptococcus salivarius
C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 15-Oct-1999
C:Accession: A44811; S22726; S28809
R:Gliffard, P.M.; Simpson, C.L.; Milward, C.P.; Jacques, N.A.
J. Gen. Microbiol. 137, 2577-2593, 1991
A:Title: Molecular characterization of a cluster of at least two glucosyltransferase ge
A:Reference number: A44811; MUID:92148377
A:Accession: A44811
A:Molecule type: DNA
A:Residues: 1-1518 <GIF>
A:Cross-references: EMBL:Z11873; NID:g47526; PIDN:CAA77900.1; PID:g47527
A:Note: sequence extracted from NCBI backbone (NCBI:81050, NCBI:P:81052)
C:Genetics:
A:Gene: gtfJ
C:Superfamily: cpl repeat homology
C:Keywords: glycosyltransferase; hexosyltransferase
F:1307-1326/Domain: cpl repeat homology <CPA>

Query Match 43.2%; Score 1231.5; DB 2; Length 1518;
Best Local Similarity 44.5%; Pred. No. 8.2e-67;
Matches 265; Conservative 93; Mismatches 168; Indels 69; Gaps 11;

hypothetical protein MYPU_2120 [imported] - Mycoplasma pulmonis (strain UAB CRIP)
C:Species: Mycoplasma pulmonis
C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 03-Aug-2001
C:Accession: D90538
R:Chambaud, I.; Hellig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galisson, F.; Moszer, I.;
Nucleic Acids Res. 29, 2145-2153, 2001
A:Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pulmonis
A:Reference number: A95112; MUID:21267165; PMID:11353084
A:Accession: D90538
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1377 <KUR>
A:Cross-references: GB:AL445566; PID:gl4089625; PIDN:CAC13385.1; GSPDB:GN00153
A:Experimental source: strain UAB CRIP
C:Genetics:
A:Gene: MYPU_2120
A:Genetic code: SGC3

Query Match 5.3%; Score 150; DB 2; Length 1377;
Best Local Similarity 20.4%; Pred. No. 0.47;
Matches 125; Conservative 78; Mismatches 208; Indels 202; Gaps 28;
QY 14 IDNSNPVQAEQLWLYLLNGFTI-----TANNDQANEDS-----VRDAPDNT---- 58
DB 629 VSSNPI--ENENLWHISKNETIKNDVLQSVKNTKYFTSDHKGVPVIAKDNVYK 685
QY 59 -----DADLNLNAQDYFNAAYGMSDAVSNKHINLEEDNADHAPDFENKI 103
DB 686 LVGNNTLLQSLADTKDIETSPKVF-LVY-----STKNRVLVLD-----SFNK- 731
QY 104 GNPQLWDDTIKSLNHLSDATRWGLDAIVHQS-----LADRENNST-----ENVVI 152
DB 732 -----TTIDHYSITSTNTSSNPPNNDQIKLHRTKIPSTTNNSSTSTNNNDV- 786
QY 153 PNFYSFVRADNNSQD-QION-----ATRDVTG-----KDYHTFPEDEKQIDAV 196
DB 787 --SSFYIYRHNDKSVSIENWEFSTKRWAKQNVTSANGDKQTYTSLNKENVKITDVF 844
QY 197 IQDNSTV-----KYNL-----YNIPASYAILTNKDTIPRVY---- 231
DB 845 IDP-STKSLMLLKGTDNSQFNQPEKQVNTILNKYNISKVNFSAEEFTQKQNYLGS 904
QY 232 GOLYTDGGYMEHQTRYDYLTLNLKSRVYVAGGSMQT--MSVGGNNILLTSRVYK 289
DB 905 GK1OTTQASWD-----YQLVTKQSTQTVINAQNIINTSYLAAGAKNALVQTNDKKV 959
QY 290 AMTATD-----TGTDETQTQIGVVVSNTPNLKGVNDKVVLLH----- 327
DB 960 YLIITGNPKYLSAGFENTKYNAGVATISDVQKKLDLSETKYSHFSDFALESIVNDG 1019
QY 328 MGAHKNQYRAAVLTDTGIVNTSDQGPAPVAMTDGDLYLSSNLVNVNGKEADTAV 387
DB 1020 FGGYLYGKTLTLRLNSQFEVINYHDFSTKMSKITVALDL--KSNIVYNGTK-KDNTI 1075
QY 388 QG---YANPDVSGVLAVVWVPGVSDAQDARTAPSTKNSGNSAYRTNAAFDSNVIFEAFS 444
DB 1076 QGVVEYLN-----LPSKGFVYKQS 1096
QY 445 NFVYTPFKESF-----RANVRIAQADFFASLGFTSFEMAPQY-----NSSKD 487
DB 1097 SNLYQPKDSEETKQDANKSEDXVYVDQKANGKNTIASQYKLDQKVFIPKNSKD 1156
QY 488 R-----TFDSTID 496
DB 1157 ALKEITPLGTIN 1169
RESULT 18
ALSN
N:alpha-amylase (EC 3.2.1.1) precursor - Bacillus amyloliquefaciens
N:Alternate names: 1,4-alpha-D-glucan glucohydrolase
C:Species: Bacillus amyloliquefaciens

C:Date: 30-Nov-1980 #sequence_revision 30-Jun-1987 #text_change 18-Jun-1999
C:Accession: A92389; A90307; I39756; I39763; A00843
R:Takkinen, K.; Pettersson, R.F.; Kalkkinen, N.; Palva, I.; Soderlund, H.; Kaarlaine,
J. Biol. Chem. 258, 1007-1013, 1983
A:Title: Amino acid sequence of alpha-amylase from Bacillus amyloliquefaciens deduced
A:Reference number: A92389; MUID:83108808
A:Contents: PUB110
A:Accession: A92389
A:Molecule type: DNA
A:Residues: 1-514 <TAK>
A:Cross-references: GB:J01542; GB:J01543; GB:M12033; GB:M12034; NID:gl42428; PIDN:AAJ
R:Chung, H.S.; Friedberg, F.
Biochem. J. 185, 387-395, 1980
A:Title: Sequence of the N-terminal half of Bacillus amyloliquefaciens alpha-amylase
A:Reference number: A90307; MUID:80241725
A:Accession: A90307
A:Molecule type: protein
A:Residues: 32-53, 'I', 55-63, 'L', 65-78, 'D', 80-83, 'S', 85-222 <CHU>
R:Palva, I.; Pettersson, R.F.; Kalkkinen, N.; Lehtoavaara, P.; Sarvas, M.; Soderlund,
Gene 15, 43-51, 1981
A:Title: Nucleotide sequence of the promoter and NH2-terminal signal peptide region
A:Reference number: I39756; MUID:83051296
A:Accession: I39756
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-96 <RES>
A:Cross-references: EMBL:V00092; NID:g39297; PIDN:CAA23430.1; PID:g39298
R:Ruohonen, L.; Hackman, P.; Lehtoavaara, P.; Knowles, J.K.C.; Karanen, S.
Gene 59, 161-170, 1987
A:Title: Efficient secretion of Bacillus amyloliquefaciens alpha-amylase cells by it:
A:Reference number: I39763; MUID:88137952
A:Accession: I39763
A:Status: Preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-39 <RE2>
A:Cross-references: GB:M18424; NID:gl42430; PIDN:AAA22192.1; PID:gl42431
C:Function:
A:Description: catalyzes the hydrolysis of internal 1,4-alpha-D-glucosidic bonds
A:Pathway: glycogen/starch degradation
C:Superfamily: alpha-amylase, amyloliquefaciens type; alpha-amylase core homology
C:Keywords: extracellular protein; glycosidase; hydrolase; polysaccharide degradation
F:1-31/Domain: signal sequence #status predicted <SIG>
F:32-514/Product: alpha-amylase #status predicted <MPT>
F:229-362/Domain: alpha-amylase core homology <AMY>
F:133,231,286/Binding site: calcium (Asn, Asp, His) #status predicted
F:162,292,359/Active site: Asp, Glu, Asp #status predicted

Query Match 5.2%; Score 149.5; DB 1; Length 514;
Best Local Similarity 22.3%; Pred. No. 0.12;
Matches 60; Conservative 35; Mismatches 103; Indels 71; Gaps 10;
QY 7 EFLANDIDNSNPVQAEQLWLYLLNGFTITANNQANFDSVRVADPNDADLNTA 66
DB 225 DYLMKADVDYDHPVVAETKKGWIYAN-----ELSDGFRIDAAKHIK---FSFL 272
QY 67 QDYFNAAGMSDAVSNKHINLEEDW-NHAD--PEYFNKIGNPQLTMDTIKNSLNHGLS 123
DB 273 RDMVQAV----ROATKEMFTVAEYWNAGKLENTLN-----KTSFNQSVF 315
QY 124 DATNRGLDAIVHQSADRENNSTENVIPNY-----SFVRAHDNNSDQIQNAIRDVTG 178
DB 316 DYPLHFNLAQASQGGYDWRLLDGTVVRHPEKAVTFVENHDT----- 360
QY 179 KDYHTFPEDEKQIDAVYQDNQSTVKKYNLYNIIPASYAILTNKDTIPRVYGLYTDG 238
DB 361 -----QPG-----QSLESTVQTWFK---PLAYAFILTRSGYPQVFGDMYGT 401
QY 239 GQYMEHQTRYDYLTLNLKSRVYVAGGQ 267
DB 402 GTSFKRIPSLKDNIEPIKARKEYAYGPQ 430

RESULT 19
AD1841
hypothetical protein alr0276 [imported] - Anabaena sp. (strain PCC 7120)
C:Species: Anabaena sp.
A:Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 11-Jan-2002
C:Accession: AD1841
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchiy
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A:Reference number: AB1807; MUID:21595385; PMID:11759840
A:Accession: AD1841
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2348 <KUR>
A:Cross-references: GB:BA000019; PID:BA77600.1; PID:g17135254; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr0276

Query Match 5.2%; Score 149.5; DB 2; Length 2348;
Best Local Similarity 21.4%; Pred. No. 1.1;
Matches 133; Conservative 82; Mismatches 233; Indels 173; Gaps 32;

QY 12 NDIDNSPVVQAEQLNWLNYLLNFGTITANNQANFDSVRVDAPDNID-ADLNLNIAQDYF 70

Db 423 NQVTSPTLFTFPQNW--NIAOTVTVTAVD-----DVIEGPKHTTIAHNVS 466

QY 71 NAAVGMDSVSNKHINILNEDWHDADPEYFNKIGNPOLTMDDTIKNSLHGLSDATN--- 127

Db 469 SSDTNYGIALANINITD--NDAPP--NTNVVQIT--EYMTGANCEFEVFNIGT 521

QY 128 -----RWGLDAIVHOSLADRENSTENVIPNVSF-----VRAHD 162

Db 522 TPVDFTCWSFDD--NTRIAGSFNLSAFGIQPGESVLTETAETETANLPTSVKII 579

QY 163 NNSQ-----DQIGNAIRDVNGKDYHTTFEEOKGIDAYIQDON-----SVVKVNLNIP 213

Db 580 NSNGGLRADEIN--LYDSTGLIDRLAYNDEV--FWGTITNTQASGWTAAANDAFEIT 635

QY 214 ASYAILTNKDTIPRVYGLDLYTGQGYMEHOTRYDILNLKSRVYVAGGOSMOTMS 273

Db 636 TNWLSAINQNSRLSTGNDVGNPGYIPNPV-----STVGAKITV-----NPSTDF 685

QY 274 VGGNNILTSVRYGKGAMTAT--DYGCTDTRFQIGVGVVSWT-----PNLKLGV---NDKV 325

Db 686 LDGQNLFPVLPQIGAGASGVINDPTDPARTILGTFINFTLSDTPVENLTITVTSSQAVV 745

QY 326 -----LHMGAHKNQYRAA-----VLTITDG-----VINYSDGAPVAMTDENG 366

Db 746 PDANILTGTCABENLKNPAGVGLANITLTVSDGTLSYIINYAASGS----- 796

QY 367 DLYLSSNLVYNGKEADTAVQGVANFDVSGYLAVVYVPGASQNDARTAPSTEKNSGNS 426

Db 797 ---VSPDTRFLTGTSDASSALAIKAN-----YMFV-----ADDEDQITRLYDRNSGLP 842

QY 427 AYRTNAFD-----SNVIFEA---PSNFVYTPTKESERANVIAQNAFFF----- 468

Db 843 L-----ASPDFTSLLGLSGSSEVDIEASTRICTIYWSGSHNSNGNDSPNRERIFATQI 898

QY 469 -----ASIGTFSFEMAPQYNSSKDRFLDSTI-----DNGYAFATDRYOLGMSPNKYGT 517

Db 899 SGTGASATLTGGYYQ-----FLEDDILANDNNHNGHGLGAGF-LGLAESAANGV 946

QY 518 DEDLNATQALHKLHAGLOVHAD 538

Db 947 SPEIRNGENI---EGLTVAPD 964

RESULT 20
T31102

filamentous hemagglutinin 1 - Haemophilus ducreyi
C:Species: Haemophilus ducreyi
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
C:Accession: T31102
R:Ward, C.K.; Lumbley, S.R.; Latimer, J.L.; Cope, L.D.; Hansen, E.J.
J. Bacteriol. 180, 6013-6022, 1998
A:Title: Haemophilus ducreyi secretes a filamentous hemagglutinin-like protein.
A:Reference number: 220984; MUID:99030326
A:Accession: T31102
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-4152 <NAR>
A:Cross-references: EMBL:AF057695; NID:g3929017; PID:g3929018; PID:AACT9757.1
C:Genetics:
A:Gene: lrpA1

Query Match 5.2%; Score 147; DB 2; Length 4152;
Best Local Similarity 19.1%; Pred. No. 3.6;
Matches 105; Conservative 85; Mismatches 181; Indels 180; Gaps 26;

QY 2 TYRGEFLANDIDNSPVVQAEQLNWLNYLLNFGTITANNQANFDSVRVDAPDNID 61

Db 593 TVNGSHVYVNGDIVSKDKAIV-----TFSNNSDFTSNGSKLVDAQNLTVN 638

QY 62 L--MNTAQD-----YFNAAVGMDSDAVSNKHINILNEDWHDADPEYFNKIG-----NPOL 108

Db 639 VNNFNITQGEIILHGNVILNAGNFTNSGNTLTKMELNISIESFINAGNLTGKNLEV 698

QY 109 THDDTTKN-----SLN-HGLSDATNR--WGLDAI-----VHOSLADRE-- 143

Db 699 HSNNTTVKNDGKLVSIENLATISSKTDTFTNGTLGLALKIAGSGNFTNAGSLASNKSL 758

QY 144 ---NKTENVVIP-----NYSFVRAHDNNSQDQIONAIRDVTGDKYHTTFEDEQ 190

Db 759 DIYGNFTNNGTIESVKSLLTNNTNFI---NNA-----TIKSYGVLTITSG 803

QY 191 KGIDAYIQQNSVVKVYNLYNIPA-----SYAIL 219

Db 804 N----FTNDSNGTVMSHDLNLTNSQANIKNKLACGGIINTAKGNITNDSNSTAI 859

QY 220 LTKWD---TIPRVY-YGDLTYDGG-----QYMEHOTRYDILNLKSRVYVAGGOS 268

Db 860 HSNNDINLANNNKVVYNGEYISQAGNISVEAKLLHNDVKLSGNITTTTTSK-----GNAT 913

QY 269 MOTMSVGGNNILTSVRYGKGAMTATDGTDETRTQIGVGVVSVNTNKLGVNDKVL 328

Db 914 VKTNSIGGGUHDANSIRVGELTLNGKFAIDLN-----OLKVALRGKI--YA 957

QY 329 GA-----AHKNQYRAAVLTITDGVINYS--DQAPVAMTDENGDIYLSHNLV 378

Db 958 GSNLTPKAKEGEKEKSTAOAKIINRGITINVKNKLEYSNVVDVNNMRSQVNL 1017

QY 379 GKEADTAVOGYANPOVSGYLAVVYVPG---ASDQDARTAPSTEKNSGNSAYRTNA 435

Db 1018 GD-----NP-----ITTLKNGVTFKDFGNRRRRASNDGEGTKKTFDNVA-- 1059

QY 436 SNVIFEAFSNF 446

Db 1060 -HLIEEAFSGY 1069

RESULT 21
S72635

exo-poly-alpha-galacturonosidase (EC 3.2.1.82) precursor - Thermoanaerobacterium ther
C:Species: Thermoanaerobacterium thermosulfurigenes
C:Date: 29-Jul-1997 #sequence_revision 29-Jul-1997 #text_change 15-Oct-1999
C:Accession: S72635; S72621

R:Matuschek, M.; Sahn, K.; Bahl, H.
submitted to the EMBL Data Library, March 1996
A:Description: Characterization of genes from Thermoanaerobacterium thermosulfurigen
A:Reference number: S72635
A:Accession: S72635

A:Molecule type: DNA
A:Residues: 1-1148 <MAT>
A:CROSS-references: EMBL:U50951; NID:G1542972; PIDN:AAB08040.1; PID:G1255234
A:Experimental source: strain EM1
R:Matusechek, M.; Sahm, K.; Zibat, A.; Bahl, H.
Mol. Gen. Genet. 252, 493-496, 1996
A:Title: Characterization of genes from *Thermoanaerobacterium thermosulfurigenes* EM1 that
A:Reference number: S72621; MUID:97033355
A:Molecule type: DNA
A:Residues: 729-1148 <MAW>
A:CROSS-references: EMBL:U50951
A:Experimental source: EM1
C:Genetics:
A:Gene: *pgla*
C:Superfamily: Thermoanaerobacterium endo-1,4-beta-xylanase A; S-layer repeat homology;
a xylanase A cellulose-binding repeat homology
C:Keywords: glycosidase; hydrolase; polysaccharide degradation
F:1-30/Domain: signal sequence #status predicted <SIG>
F:31-1148/Product: exo-poly-alpha-galacturonidase #status predicted <MAT>
F:969-1022/Domain: S-layer repeat homology <SLR1>
F:1028-1081/Domain: S-layer repeat homology <SLR2>
F:1092-1144/Domain: S-layer repeat homology <SLR3>
Query Match 5.1%; Score 145.5; DB 2; Length 1148;
Best Local Similarity 19.0%; Pred. No. 0.67;
Matches 139; Conservative 89; Mismatches 252; Indels 253; Gaps 35;
QY 3 YRGEFFLLAN-----DIDNSNPVQAEQLNWLYYL-----LNFGRIT-----AN 41
DB 368 YIGGLIVNPSPHTTSVSQSNQVNLQLIASYDCNNGGIDFGNGSTGLTVVNSVFNIG 427
QY 42 NQANFDS-VRVDPONIDADLMNIADQYFNAAAGHSDAVSNKHI-----NILEDNHHA 95
DB 428 DDVNFEDAGVGLSGEQNPPTGNAWFDNYFGRGHV---IANGSHRAAMQIQLAEDNVI 484
QY 96 DPEVFNKIGNPQ-----LPMDDTIKNSLNHGLSDA-----TNKGLD 132
DB 485 NGTAIGLGRKSQSGNGGGARNITFROSALAYITDNDGSPFLDTGYSSALPTDTSNAPD 544
QY 133 -----AIYHQSLAD-----RENNSTE 148
DB 545 EPTFHDITVENCIVNGSKKYAIFMGAPGDFNIFNVPVFGAGYQTKYIYKLSFNF 604
QY 149 NVVI-----PNYS-----FVRAHD-----NNSQDI-----QNAIRDVTGKDYH-- 182
DB 605 NVFYGSTPNYDGTTPPNPWHFVHSTNIQFTGNTTPQPIPEWPSNSTATVSNVTYDSA 664
QY 183 TTFEEDQKGDAYIQDQNSTVKKYNYLNIPASYAILLTNKTPIPVYIGDL-----234
DB 665 TVTWISAVSG-----TYTGQVLGADGTILGKVDASTTYY--DLGLKLPDQTQ 709
QY 235 YDGGQYMEHQTR--YPTLTNLKLSRVKYYVAGGQSMGTSMVGGNNNLT--SVRYGKGAMT 292
DB 710 YTTVTEATDQNNITFGPIVTFITAKAS-SAGQGGSTGGTNGSGNTGSANTSSGIN 768
QY 293 ATDGTDETRTQGIQGVVSWTPNKL-----GVNDKVVLUHGAHKNQOYRA 339
DB 769 TTTNNNGTSSVTGTTKNGNVTLLILDAGKADIIANSKDKKVVFDLTITIGSQG---825
QY 340 AVLTDTGVIYNTSDQAPVAMTDENGDLYS---SHNLVNGKEADYAVQYANPDV 395
DB 826 KVVQLSKDILTSAANGKDIVKSDNASIALTKDALNQNIQNG---VNVSTKNGKPNV 882
QY 396 SGYLAIVPVVPGASDQADAPSPPEKNSGNSAYRTNAAFDSNVIFAFSN-----FVY 448
DB 883 TNY-----VLSNVVDITISGS-----SGNVALAKPVEVTLNI---SKANDPKVAYVY 929
QY 449 TPTKES-----ERANVRIQAQNADEFASIGFTSFEMAFQYNSKD-----R 489
DB 930 NPTTNQMEYVGKVDASSGTTTFNTHFSQ---YAAFEYDKTFNDIKDNWAKDVIETLASR 987

QY 489 TFLDSTIDNGVAFTRDYDLGMSEPNK-----YGTDEDLR-----522
DB 988 HIVEGMDTQY-----EPNKTVTTRAEFTAMILLRLNIKEEAYSDFSYKSGDW 1036
QY 523 --NAIQALHKAGL 533
DB 1037 YANAIEAYKAGI 1049
A:Molecule type: DNA
A:Residues: 1-518 <TSU>
A:CROSS-references: GB:M18862; NID:G142496; PIDN:AAA22231.1; PID:G142497
A:Experimental source: chromosomal DNA of strain 707
A:Note: amino end of mature protein also determined
C:Comment: this is the smallest of five starch-hydrolyzing enzymes from this organism
C:Function:
A:Description: catalyzes the hydrolysis of internal 1,4-alpha-D-glucosidic bonds
A:Pathway: glycogen/starch degradation
C:Superfamily: alpha-amylase, amyloliquefaciens type; alpha-amylase core homology
C:Keywords: extracellular protein; glycosidase; hydrolase; polysaccharide degradation
F:1-33/Domain: signal sequence #status predicted <SIG>
F:34-518/Product: alpha-amylase #status experimental <MAT>
F:236-369/Domain: alpha-amylase core homology <AMY>
F:139,238,273/Binding site: calcium (Asn, Asp, His) #status predicted
F:269,299,366/Active site: Asp, Glu, Asp #status predicted

Query Match 5.1%; Score 144.5; DB 1; Length 518;
Best Local Similarity 20.7%; Pred. No. 0.24;
Matches 74; Conservative 56; Mismatches 113; Indels 115; Gaps 16;

QY 7 EFLANDIDNSNPVQAEQLNW-LYLLNFGITANDQANFDSVRVDAPDNDADLMNI 63
DB 232 DYLMYADIDMDHPVVEVNELRNGVVTNTLG-----LDGFRIDAVKHK---YSF 278
QY 66 AQDYFN---AAYGMDSDAVSNKHINILEDNHNADPEYFNKIGNPQLTMDDTIKNSLNHGL 122
DB 279 TRDWINHVSATCKNMFVA-----EFNK-----NDLGALENYLQ---KTNWNSV 321
QY 123 SDATNRWGLDAIVHQSL--ADRENNSTENVVPNYSFVRAHDNNSQDIQNAIRDVTGKD 180
DB 322 -----FDVPLHLYNLNASKSGGNYDMRNIFNGTVVQRHPS-----356
QY 181 YHTTFEDE-----QKGDYATQDQNSTVKKYNYLNIPASYAILLTNKTPIPVYIGDLY 235
DB 357 -HAVTVFDNDHDSQPEALSFVEE-----WFKPLAYALTIREQGYFSVFFGDYI 405
QY 236 ---TDGQYMEHQTRYDTLTNLLKSRVYVAG-----QQSMQTWASVGGNNHI 280
DB 406 GIPTHGVPMARSK-----IDPILEARQKYAYGKQNDYLDHNNIIGWREGNTAHPNSGL 459
QY 281 LTVRYVKGKA-----MTATDGTDETRTQGIQGVVSWTPNKLGVN 321
DB 460 ATIMSDGAGGSKWFWGRNKAGQVSDITGNRTGTVTINADGNGNFSVNGSGVSIWVN 517

RESULT 23
E89921
hypothetical protein ebha [imported] - *Staphylococcus aureus* (strain N315)
C:Species: *Staphylococcus aureus*
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 23-Oct-2001

C:Accession: B89921
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Ogura, A.; Mizutani-Oi, Y.; Kobayashi, N.; Sawano, R.; Inoue, R.; Kaito, C.; Sekimizu, K.; C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A:Title: Whole genome sequencing of metacillin-resistant *Staphylococcus aureus*.
A:Reference number: A89756; MUID:21311952; PMID:11418146
A:Accession: B89921
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-6713 <KUR>
A:Cross-references: GB:BA00018; PID:g13701232; PIDN:BA842527.1; GSPDB:GN00149
A:Experimental source: strain N315
C:Genetics:
A:Gene: ebhA

Query Match 5.0%; Score 144; DB 2; Length 6713;
Best Local Similarity 19.8%; Pred. No. 11;
Matches 103; Conservative 68; Mismatches 193; Indels 156; Gaps 23;

QY 67 QDYFNAAYGMDSDAVSNKHINILEDWHDPE-----YFNKIGNP-----Q 107
Db 247 QSLNNAKGLRDSIANETTVKASQNYTASPNQSTYNSAVSNAGIINQTNPTMDTSA 306
QY 108 LTMDDTIKSLHGLSDATNRWGLDAIVHQSLADRENSTENVVPIYSEVRAHDNNSQ- 166
Db 307 ITQATTVQVNAKNGLNGAEN-----LRNAQNTAKQNL-----NTLSHLTNQK 349
QY 167 ---DQTONA---TRDVTGDKYHTFTFDEQKGDVAYIQDQNSTVKKYNYLNPASVAILL 220
Db 350 SAISSQIDRAGHSEVTAAKNAATELNAQNGLEQAHDQNTVKQ-----GVNF 398
QY 221 TNKDTIPRYVYGLDYGQYME---HQRYVDTLNLKSRVYVAGQSMQSMVSGN 277
Db 399 TDADKAKR---DAYTNVSRATILNKTKGTANTSKQDVEAAIQNTSAKN-----ALNGD 450
QY 278 NNILTSVRYGKGAMT-----ATDGTDETFQIGVGVVSNY-----PNKL 318
Db 451 QNYTNAKNAKNAALNLTSTINNAKRDATTKIDQAVTVACVEAVSNTGTQLANTAMANLQ 510
QY 319 GVNDKV-----VLMGAAHKNQYRAAVLTITDGTGIVNYTSDQAPVAMTDE----- 364
Db 511 GINDKANTLASENHDADSKTAYTQAV-TNAENILKNKSGNLDKAAVENALSOVTNA 569
QY 365 ---NGDLYL---SSRNIVYNGKEADTA-----VQGYANPD----- 394
Db 570 KGALGNHNLQAKSNANTYINGLQHLTFAQDKLQVQQAQNVAGVDTVKSSANTLNG 629
QY 395 VSGYLAVVPGVGSNDQDARTAPSTEKNSNSAYRNAFDSNVIFEAFSNFYVPTKES 454
Db 630 AMGLRNSIODNTATKNGQNYLDATERNKTN---YNNVADSANGVINATSN-----PNMDA 682
QY 455 ERANVRIAQNADFASLGTSTSEMAPQYNSKDKTFLDST 494
Db 683 NAIN-----QIATQVSTKNA---LDGT 702

RESULT 24
G71676
hypothetical protein RP226 - Rickettsia prowazekii
C:Species: Rickettsia prowazekii
C:Date: 21-Nov-1998 #sequence_revision 21-Nov-1998 #text_change 03-Nov-2000
C:Accession: G71676
R:Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sichteritz-Ponten, T.; Alsmark, Nature 396, 133-140, 1998
A:Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.
A:Reference number: A71630; MUID:99039499
A:Accession: G71676
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-584 <ND>
A:Cross-references: GB:AJ235271; GB:AJ235269; NID:g3868717; PIDN:CAA14689.1; PID:g386078

A:Experimental source: strain Madrid E
C:Genetics:
A:Gene: RP226

Query Match 5.0%; Score 143; DB 2; Length 584;
Best Local Similarity 19.2%; Pred. No. 0.36;
Matches 103; Conservative 74; Mismatches 165; Indels 194; Gaps 23;

QY 62 LMNIAQDYFNAAYGMDSDAVSNKHINILEDWN-----HADPEYFNK---IGNPOLT 109
Db 109 LPSVASNY-----IDSKAEVENSTNLAASYNQDIQVKQBFDPSEASEPTDIGNTKFT 162
QY 110 HDDTIKSLHGLSDATNRWGLDAIVHQSLADRENSTENVVPIN-----YSFVRAHDNN 164
Db 163 -----SATNHEMYK-----EAV-----SSNDKETNLTSNIITPNVPSVSIPTAQDVN 206
QY 165 ----SDDQIQ-----NAIRDVTGDKYHTFTFDEQKGDVAYIQDQNSTVKKYNYLNP 213
Db 207 YVVPQSQSVQIYKPTNLASIRNPILNHH-----DLNKKVEK---NLESTI 249
QY 214 ASYAILLNNKDIPIPKVYVYGLDYGQYMEHQRYVDTLNLKSRVYVAGQSMQSM 273
Db 250 SNTTIPITNWSVPSI-----QDTITQTLNITVPTAETHSVQST 290
QY 274 VGGNN-----NILTSVRYGKGAMTATDGTDETFQIGVGVVSN-----TPN 315
Db 291 VMSNOHSQAQIPISINPTVETSTVLRATESVPINNQSQEIFVSESESTKKQDMYTP 350
QY 316 LKLGIV-----NDKVLHMGAAHKNQYRAAVLTITDGTGIVNYTSDQGA 357
Db 351 MPVLVYDVPNKSQSKPLALBQKNNQDILNNAQESHSVSSSVNTIQKQNDKYNNAIS-EST 409
QY 358 PVAMTDENGDLXLSNHLVYNGKEADTAQGYANPDVSGYLAV-----WYPVGASQ 411
Db 410 KEFVKNETQMLPDDDIVL-GKLTEDATLE---QMDHGYIKLFQKKEWI----- 457
QY 412 DARTAPSTEKNSNSAYRNAFDSNVIFEAFSNFYVPTKESERANVRIAQNADFASL 471
Db 458 -----ANAKRKKVSEFIKY-----DDINKNKDIYANL 486
QY 472 GFTS-----FEMAP-----VYNSSKDKTFLDSTIDNGYAFDTR 505
Db 487 SYISAVDNNAFRAVEKNLFEALRDLDVYILQAKNSTGETILTSSYNGNYILAKF 542

RESULT 25

ALBSCG

N:Alternates: cyclomaltoextrin glucanotransferase (EC 2.4.1.19), precursor [validated] - Bacillus
C:Species: bacillus circulans
A:Variety: strain 8
C:Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 15-Sep-2000
C:Accession: S23674; S23772
R:Nitschke, L.; Heeger, K.; Bender, H.; Schulz, G.E.
Appl. Microbiol. Biotechnol. 33, 542-546, 1990
A:Title: Molecular cloning, nucleotide sequence and expression in *Escherichia coli*
A:Reference number: S23674; MUID:91103970
A:Accession: S23674
A:Molecule type: DNA
A:Residues: 1-718 <NIT>
A:Cross-references: EMBL:X68326; NID:g39419; PIDN:CAA48401.1; PID:g39420
R:Schlitz, E.; Hofmann, B.
submitted to the Protein Sequence Database, September 1992
A:Reference number: S23772
A:Accession: S23772
A:Molecule type: protein
A:Residues: 35-49; 'X', 167-173, 'X', 175-180 <SCH>
R:Klein, C.; Schulz, G.E.
submitted to the Brookhaven Protein Data Bank, June 1992
A:Reference number: A51664; PDB:ICGT
A:Contents: annotation; X-ray crystallography, 2.0 angstroms, residues 35-718
R:Klein, C.; Schulz, G.E.

J. Mol. Biol. 217, 737-750, 1991
A:Title: Structure of cyclodextrin glucosyltransferase refined at 2.0 A resolution.
A:Reference number: S14381; MUID:91171298
A:Contents: annotation; X-ray crystallography, 2.0 angstroms
C:Function:
A:Description: catalyzes the breaking and reformation of 1,4-alpha-D-glucopyranosyl bond
A:Note: can also catalyze the disproportionation transfer of 6 to 8 1,4-alpha-D-glucopyranosyl
C:Superfamily: cyclomaltodextrin glucanotransferase; alpha-amylose core homology
C:Keywords: calcium; duplication; extracellular protein; glycosyltransferase; hexosyltra
F:1-34/Domain: signal sequence #status predicted <SIG>
F:35-718/Product: cyclomaltodextrin glucanotransferase #status experimental <WAT>
F:35-172/Domain: A1 <DA1>
F:173-236/Domain: B <DOB>
F:231-365/Domain: alpha-amylose core homology <AMY>
F:237-440/Domain: A2 <DA2>
F:441-528/Domain: C <DOC>
F:529-614/Domain: D <DOD>
F:615-718/Domain: E <DOE>
F:61-63,66,67,85,87/Binding site: calcium (Asp, Asn, Asn, Gly, Asp) #status experimental
F:77-84/Disulfide bonds: #status experimental
F:173,224,233,267/Binding site: calcium (Asn, Ile, Asp, His) #status experimental
F:263,291,362/Active site: Asp, Glu, Asp #status experimental

Query Match 5.0%; Score 141.5; DB 1; Length 718;
Best Local Similarity 21.8%; Fred.No. 0.59;
Matches 105; Conservative 55; Mismatches 184; Indels 137; Gaps 22;

QY 64 NTAQDYFAAYGMSDAVSN---KHINILEDWHDAPYFNKIGNPQLTMDTIKNSLNH 120
DB 203 NTNGYFHNGSDFSLENGIYKNLYDLADFNHN-----ATIDKVFDAIKL 251
QY 121 GLSDATNRWGLDAIVHQSLADREN-----NSTENVTIPNYSF 157
DB 252 WLDMGVDGIRVDVAVKHMPGLGOKSWMSIYAHKPVTFGENFLGSAADADNTDFANKSG 311
QY 158 VRAHDNNSODIQNAIRDYTKDYHTFTFEDEQKIDAVIOPQNSTVKKYNLYN----- 211
DB 312 MSLLDFRNSAVRNVFRONTSMY-----ALDSMI-----NSTADYKQVNDQVTFI 358
QY 212 -----IPASVAILLTNRKDTIPRVYVG-DLYTDGGQYMEHQTR----- 247
DB 359 DNHDMDFRXTSAVNNRRLQALAFILTSRG-VPAYIYGTEQYLTGNGDPDNRKMPFSFK 417
QY 248 -----YYDTLNLKSRVYVAG-----GQSHQTMVSGGNHILTS 283
DB 418 STTAFNVISKLAPLRKSNPAIAYGSTQQRWINNDVYVYERKFKSVAVVAV--NRHLSTS 475
QY 284 VRYGKAMTATDTGDTERTOGIGVYVVSNTPLKLGNDKVVHLHMGAAHKNOQYRAAVLT 343
DB 476 ASI-TGLSTSLPTGS-YTDVLG-GVLNGNITSTNGSINNFTLAAGAT-AVWQYTTAETT 531
QY 344 TTDGVINYTSQQGAPVAMTDENGDLVLSHNLVYNGKEADTAVQGYANPOVSGYLAVWV 403
DB 532 PTIGHVGPVMGKPGNVVTIDGRG--FGSTKGTVIFG---TTAVTGAA-----ITSW- 577
QY 404 PVGASNDQARTAPSTKNSGSAAYTNAAFDSNVIFEAFSNFVYPTTKESERANVRLAQ 463
DB 578 -----EDQIKVTIPSA--AGNYAVKVA--SGVNSNAYNNFTIL---TGDQVTVRFVW 625
QY 464 N 464
DB 626 N 626

Search completed: August 12, 2002, 08:00:17
Job time: 249 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 12, 2002, 08:00:23 ; Search time 13.48 Seconds

(without alignments)
1559.697 Million cell updates/sec

Title: US-09-995-749A-2_COPY_972_1514

Perfect score: 2852

Sequence: 1 GTYRGQSELLANDINSNFV.....AIQALHKAGLOYMADWVDPQ 543

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1464.5	51.3	1375	1	GTFC_STRMU
2	1451.5	50.9	1476	1	GTFB_STRMU
3	1413	49.5	1462	1	GTFF_STRMU
4	1405	49.3	1592	1	GTFF2_STRDO
5	1405	49.3	1597	1	GTFF1_STRDO
6	1219.5	42.8	1365	1	GTFS_STRDO
7	156.5	5.5	1861	1	APU_THETU
8	155.5	5.5	512	1	AMY_BACLI
9	149.5	5.2	514	1	AMT6_BACAM
10	144.5	5.1	518	1	AMT6_BAC7
11	141.5	5.0	718	1	CDGT_BACCI
12	141	4.9	495	1	AMY2_ECOLI
13	139.5	4.9	713	1	CDGU_BACCI
14	137	4.8	566	1	NPRE_BACCE
15	137	4.8	1902	1	P2P_LACPA
16	136	4.8	1026	1	VG37_BPT4
17	134	4.7	718	1	CDGT_BACSS
18	132.5	4.6	828	1	PMFC_PROMI
19	131.5	4.6	1104	1	COLA_CLOPE
20	131	4.6	1279	1	APU_THESA
21	130.5	4.6	793	1	YF06_MYCPN
22	129	4.5	581	1	AMV1_SCHPO
23	129	4.5	852	1	CSG_HALHA
24	128.5	4.5	1475	1	APU_THETV
25	128	4.5	549	1	AMY_BACST
26	128	4.5	731	1	GLGB_MYCTU
27	128	4.5	1902	1	P1P_LACLC
28	128	4.5	1953	1	BIGA_SALTY
29	128	4.5	2334	1	WAPA_BACSU
30	127.5	4.5	712	1	CDGT_BACS3
31	127	4.5	2003	1	YDBA_ECOLI
32	127	4.5	2366	1	TOXB_CLODI
33	126.5	4.4	713	1	AMVR_BACS8

RESULT 1

ALIGNMENTS

34	126	4.4	711	1	CDGT_BACST
35	125	4.4	429	1	AMY6_HORVU
36	125	4.4	1300	1	120K_RICRI
37	125	4.4	1654	1	OMP_RICRI
38	124.5	4.4	713	1	CDGT_BACSP
39	124.5	4.4	1211	1	ATH1_YEAST
40	123.5	4.3	497	1	FLIC_ECOLI
41	123	4.3	1902	1	P2P_LACLC
42	122.5	4.3	765	1	Y260_MYCGE
43	122.5	4.3	1035	1	NANA_STRPN
44	122	4.3	489	1	TACP_VIBCH
45	122	4.3	1296	1	VACI_HELPV
46	121.5	4.3	435	1	AM3D_ORYSA
47	121	4.2	494	1	AMY2_SALTY
48	121	4.2	918	1	YMB_CABEL
49	120.5	4.2	928	1	PM10_CHLPN
50	119.5	4.2	905	1	HXA1_HAEIN
51	119	4.2	1655	1	OMP_RICCN
52	118.5	4.2	1226	1	METH_SALTY
53	118	4.1	1902	1	P3P_LACLC
54	117.5	4.1	504	1	FLIC_SALMO
55	117.5	4.1	575	1	FLS2_CAMJE
56	117.5	4.1	728	1	GLGB_ECOLI
57	117.5	4.1	1226	1	METH_ECOLI
58	116.5	4.1	427	1	AMY2_HORVU
59	116.5	4.1	443	1	AM2A_ORYSA
60	116.5	4.1	445	1	AMC2_ORYSA
61	116.5	4.1	504	1	FLIC_SALSE
62	116.5	4.1	574	1	FLB3_CAMJE
63	116.5	4.1	731	1	BAF1_YEAST
64	116.5	4.1	1409	1	HAP1_HAEIN
65	116	4.1	486	1	YAIT_ECOLI
66	116	4.1	507	1	FLIC_SALBE
67	116	4.1	507	1	FLIC_SALON
68	115.5	4.0	778	1	SEA2_MYCLE
69	115.5	4.0	1243	1	VG37_BPK3
70	115.5	4.0	1643	1	OMP_RICPR
71	115	4.0	673	1	FVUA_YEREN
72	115	4.0	704	1	CDGT_BACOH
73	115	4.0	939	1	EAE_ECO27
74	114	4.0	673	1	FVUA_YERPE
75	114	4.0	1122	1	ADP1_MYCGA
76	114	4.0	1577	1	HLVA_PROMI
77	113.5	4.0	504	1	FLIC_SALBU
78	113.5	4.0	575	1	FLA2_CAMJE
79	113.5	4.0	746	1	FEPA_ECOLI
80	113.5	4.0	836	1	UMEG_YEAST
81	113	4.0	571	1	FLB1_CAMJE
82	113	4.0	703	1	CDGT_BACS2
83	113	4.0	771	1	YKQ_YEAST
84	113	4.0	934	1	Y321_MYCGE
85	113	4.0	1385	1	C5AA_BACUD
86	113	4.0	1545	1	IGA3_HAEIN
87	112.5	3.9	1014	1	HEX2_YEAST
88	112.5	3.9	1017	1	VATA_PYRAB
89	112	3.9	574	1	FLA3_CAMJE
90	112	3.9	614	1	DNAC_CIOAB
91	112	3.9	650	1	SAGL_YEAST
92	112	3.9	713	1	CDGT_BACSO
93	112	3.9	718	1	CDGT_BACLI
94	112	3.9	1165	1	YNF4_YEAST
95	111.5	3.9	368	1	AMY3_HORVU
96	111.5	3.9	467	1	YATL_ECOLI
97	111.5	3.9	504	1	FLIC_ECOLI
98	111.5	3.9	861	1	YEST_HAEIN
99	111.5	3.9	1177	1	Y307_MYCGE
100	111.5	3.9	1196	1	AMVB_PAEPO

P31797	bacillus st
P04750	hordeum vul
P14914	ricketsia
Q53047	r outer mem
P30921	bacillus sp
P48016	saccharomyc
P04949	escherichia
P15293	lactococcus
P47502	mycoplasma
Q59559	streptococc
P29481	vibrio chol
Q48247	helicobacte
P27933	oryza sativ
P26613	salmonella
P34487	caenorhabdi
Q9R665	chlamydia p
P44602	haemophilus
Q9Kka3	r outer mem
P37586	salmonella
P15292	lactococcus
Q08973	salmonella
P22252	campylobact
P07762	escherichia
P13009	escherichia
P04063	hordeum vul
P27935	oryza sativ
P27941	oryza sativ
Q08983	salmonella
Q46114	campylobact
P14164	saccharomyc
P44596	haemophilus
P77199	escherichia
Q08968	salmonella
Q08974	salmonella
Q39222	mycobacteri
Q38394	bacterioph
Q53020	r outer mem
P46360	versinia en
P27036	bacillus oh
P19609	escherichia
P46359	versinia pe
Q49379	mycoplasma
P16466	proteus mir
Q06969	salmonella
P22251	campylobact
P05825	escherichia
P39001	saccharomyc
P31746	bacillus sp
Q02208	saccharomyc
P47563	mycoplasma
Q45760	bacillus th
P45385	haemophilus
Q00816	saccharomyc
Q0907	pyrococcc
Q46113	campylobact
P30721	clostridium
P20840	saccharomyc
P05618	bacillus sp
P14014	bacillus li
P53950	saccharomyc
P04747	hordeum vul
Q05981	escherichia
Q05981	salmonella
P47549	mycoplasma
P21543	paenibacill

```
GTFC_STRMU
ID GTFC_STRMU STANDARD; PRT: 1375 AA.
AC P13470; P05427;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Glucosyltransferase-SI precursor (EC 2.4.1.5) (GTF-SI)
DE (dextranase) (Sucrose 6-glucosyltransferase).
GN GTFC.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GS-5;
RX MEDLINE=89137980; PubMed=2976010;
RA Ueda S., Shiroza T., Kuramitsu H.K.;
RT "Sequence analysis of the glfC gene from Streptococcus mutans GS-5.";
RL gene 69:101-109(1988).
RN [2]
RP SEQUENCE OF 1-349 FROM N.A.
RC STRAIN=GS-5;
RX MEDLINE=87308013; PubMed=3040685;
RA Shiroza T., Ueda S., Kuramitsu H.K.;
RT "Sequence analysis of the glfB gene from Streptococcus mutans.";
RL J. Bacteriol. 169:4263-4270(1987).
CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT
CC TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE
CC OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIANE THE
CC AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.
CC -1- CATALYTIC ACTIVITY: Sucrose + ((1,6)-alpha-D-glucosyl)(N) - D-
CC fructose + ((1,6)-alpha-D-glucosyl)(N1).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- DISEASE: DENTAL CARIES.
CC -1- MISCELLANEOUS: GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA
CC 1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES BOTH
CC WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH
CC FORMS OF GLUCANS.
CC -1- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-
CC BINDING PROTEIN FROM S.MUTANS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: M22054; AAA88592.1; -;
DR EMBL: M17361; AAA88589.1; -;
DR PIR: J0345; J0345.
DR PIR: C33135; C33135.
DR InterPro: IPR002479; CW_binding.
DR InterPro: IPR003318; Glyco_hydro_70.
DR Pfam: PF01473; CW_binding.1; 7.
DR Pfam: PF02324; Glyco_hydro_70; 1.
KW Transferase; Glycosyltransferase; Signal; Repeat; Dental caries.
FT SIGNAL 1 34
FT CHAIN 35 1375 GLUCOSYLTRANSFERASE-SI.
FT DOMAIN 35 1050 CATALYTIC (APPROXIMATE).
FT DOMAIN 1126 1375 GLUCAN-BINDING (APPROXIMATE).
FT DOMAIN 1126 1375 2.4 A, 1 C AND 1 AC REPEATS.
FT REPEAT 1126 1159 A REPEAT.
FT REPEAT 1169 1200 A REPEAT.
FT REPEAT 1227 1238 C REPEAT.
FT REPEAT 1253 1303 AC REPEAT.
FT REPEAT 1318 1330 A REPEAT (INCOMPLETE).
SQ SEQUENCE 1375 AA; 153022 MW; D4B80CBEE0ACE13 CRC64;
```

Query Match

51.3%; Score 1464.5; DB 1; Length 1375;

```
Best Local Similarity 52.9%; pred. No. 3.9e-77;
Matches 292; Conservative 94; Mismatches 141; Indels 25; Gaps 8;

QY 1 QYRGGEFLANDIDNSNPVQAEOLNMLYLLNPGTTTANNDQANFOSRVDPDNIDA 60
: : |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 425 RTIGYGEFLANDVDSNPVQAEOLNMLHFLNPGNTIYANDPDANFOSIRVDAVDYDA 484
: : |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 61 DLMNTAQDYFNAAYGM-DSDAVSNKHINILEDNHDAEYFNKIGNPOLTMDDTIKNSLN 119
||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||
DB 485 DLUQTAGDYLKRAKGIHKNDKAAHDLSLLEAWSYDTPYLDHGDNDNMNDRSLRL 544
: : |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 120 HGLSDATN-RWGLDAIVHQSADRENNSNVVINYSFVRAHNNNSQDIONAIR---- 174
: : ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||
DB 545 YSLAKFLMORSCHNPLITNSLVNRTDDNNAETAAVSYSEIRAHDSVQDLNIRIETIN 604
: : |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 175 -DVTGKYHTFTFEDEQKIDYIQDQNSTVKKYMLNIPASAYALLNKKDTPRVIYGD 233
: : ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||
DB 605 PNWVG---YSFTTEIKKAFETYNKDLATEKKYTHYTALSYALLLANKSSVPRVIYGD 661
: : |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 234 LYTDGGOYMEHOTRYVDLTNLLKSRVYVAGSQSOTMSVGGNNILTSYRYGKGAMTA 293
: : |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 662 MFTDDGOYMAHKTINTEJETLTKARIKYVSGGQAWRNOQV-GNSEIITSVRYGKALKA 720
: : |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 294 TDTGDTETRTQIGYVVVSNTPNLKLGVDNKVYLHMGAAHKNQYRAAVLTITDGVINITS 353
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 721 TDTGDTTTRTSGVAVIEGNNPSSLRLKASDRVVVNMGAHKNQAYRPLLTITDNGIKAYHS 780
: : |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 354 DGAP--VAMTDENGDLVSSHNLVVNGKEADTAVQGYANPDVSGYLAIVVPGASNDQ 411
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 781 DQEAAGLVRYTDRGELIFTA-----AD--IKGYANPQVSGYLGVPVPGAAADQ 828
: : |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 412 DARTAPSTEKNKSAYRDNAAAFDSNVIFEAPSNFVYPTKESERANVRIACONADFFASL 471
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 829 DNVAASTAPSTDCKSVHQNAALDSRVMEGFSNFQAFATKKEEYTNVYAKNVKPFABW 888
: : |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 472 GFTSFEMAPQYNSSKDRDTELDSTONGYAFTRYDGMSEPKNKYGTDEDLRAIALHKA 531
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 889 GVTDFEMAPQYSSVTSGLDSVLIQNGYAFTRYDGLGSKPNKYGTADDLVRAIKALHKS 948
: : |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 532 GLOYMADWVPDQ 543
: : |||||:|||||
DB 949 GIKVMADWVPDQ 960

RESULT 2
GTFC_STRMU
ID GTFC_STRMU STANDARD; PRT: 1476 AA.
AC P08987; O69381; O69384; O69387; O69390; O69396;
DT 01-NOV-1988 (Rel. 09, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Glucosyltransferase-1 precursor (EC 2.4.1.5) (GTF-I) (Dextranucrase)
DE (Sucrose 6-glucosyltransferase).
GN GTFC.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GS-5;
RX MEDLINE=87308013; PubMed=3040685;
RA Shiroza T., Ueda S., Kuramitsu H.K.;
RT "Sequence analysis of the glfB gene from Streptococcus mutans.";
RL J. Bacteriol. 169:4263-4270(1987).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MT4239, MT4245, MT4251, MT4467, AND MT8148;
RX MEDLINE=98231643; PubMed=9570124;
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,
RA Kimura S., Hamada S.;
RT "Molecular analyses of glucosyltransferase genes among strains of
RT Streptococcus mutans.";
```

RL FEMS Microbiol. Lett. 161:331-336(1998).

CC !- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT

CC TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE

CC OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIANE THE

CC AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.

CC !- CATALYTIC ACTIVITY: Sucrose + ((1,6)-alpha-D-glucosyl)(N) - D-

CC fructose + ((1,6)-alpha-D-glucosyl)(n+1).

CC !- SUBCELLULAR LOCATION: Secreted.

CC !- DISEASE: DENTAL CARIES.

CC !- MISCELLANEOUS: GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA

CC 1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES). GTF-S SYNTHESIZES

CC WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH

CC FORMS OF GLUCANS.

CC !- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-

CC BINDING PROTEIN FROM S.MUTANS.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; M17361; AAA88588.1; -

DR EMBL; D88651; BAA26101.1; -

DR EMBL; D88654; BAA26105.1; -

DR EMBL; D88657; BAA26109.1; -

DR EMBL; D88660; BAA26113.1; -

DR EMBL; D89977; BAA26119.1; -

DR PIR; B33135; B33135.1; -

DR InterPro; IPR002479; CW_Binding.

DR InterPro; IPR003318; Glyco_hydro_70.

DR Pfam; PF01473; CW_Binding_1; 13.

DR Pfam; PF02324; Glyco_hydro_70; 1.

KW Transferase; Glycosyltransferase; Signal; Repeat; Dental caries.

FT SIGNAL 1

FT CHAIN 35 1476

FT DOMAIN 35 1051

FT DOMAIN 1097 1476

FT REPEAT 1097 1130

FT DOMAIN 1161 1470

FT REPEAT 1161 1210

FT REPEAT 1225 1275

FT REPEAT 1290 1340

FT REPEAT 1355 1405

FT REPEAT 1420 1470

FT VARIANT 62 65

FT VARIANT 65 65

FT VARIANT 68 68

FT VARIANT 78 78

FT VARIANT 86 86

FT VARIANT 89 89

FT VARIANT 168 168

FT VARIANT 276 276

FT VARIANT 399 399

FT VARIANT 474 474

FT VARIANT 512 512

FT VARIANT 519 519

FT VARIANT 701 701

FT VARIANT 708 708

FT VARIANT 938 938

FT VARIANT 952 957

FT VARIANT 963 964

FT VARIANT 968 970

FT VARIANT 1086 1086

FT VARIANT 1158 1158

FT VARIANT 1163 1163

FT VARIANT 1163 1163

FT VARIANT 1168 1168

FT VARIANT 1182 1182

FT VARIANT 1234 1234

FT VARIANT 1263 1263

FT VARIANT 1263 1263

FT VARIANT 1264 1264

FT VARIANT 1272 1272

FT VARIANT 1329 1329

FT VARIANT 1394 1394

FT VARIANT 1402 1402

FT VARIANT 1459 1459

FT CONFLICT 570 570

FT CONFLICT 800 817

FT CONFLICT 1310 1310

FT CONFLICT 1476 AA; 165685 MW; 3479862B07694D98 CRC64;

FT SEQUENCE 1476 AA; 165685 MW; 3479862B07694D98 CRC64;

Query Match 50.9%; Score 1451.5; DB 1; Length 1476;

Best Local Similarity 52.6%; Pred. No. 2.4e-76;

Matches 290; Conservative 93; Mismatches 143; Indels 25; Gaps 8;

QY 2 TYRGGFLLANDIDNSPVVQAEQLNWLIVLLNFGTITANDQANFDSVVDADPNIDAD 61

DB 400 TGGYEFLANDVDNSPVVQAEQLNWLHFLMNFNYANDQANFDSLRVDAVDVDAD 459

QY 62 LMNLAQDYFNAAYGM-DSDAVSNKHINILEDNHADPEYFNKIGNPOLMTDMTIKNSLNH 120

DB 460 LLQIAGDYLKAAGIHKNDKAANDHLSILEANSNDPTPLHDDGDMNIMNDKRLSLF 519

QY 121 GLSDATN-RWGLDAIVHQSLADRENNSTENVIPNYSFVRAHNNNSQDQIQNAIR----- 174

DB 520 SLAKPLNQRSGMPLITNSLVNRTDDNAETAAVPYSYFRAHDSVQDLIRDIKAEINP 579

QY 175 DVYTGKDYHFTFEDEOKGIDAYTQDQNSTVKKYNLYNPASVAILITNKTDTIPRVYGD 234

DB 580 NVVG---YSFTHKEIKKAEIYNKDLATEKKYTHYNTALSYALLTNSVPRVYGD 636

QY 235 YTDGGYMEHQTRYDTLNLKLSRVKVVAGQSMQTSVGGNNILTSVRYGKGAMTAT 294

DB 637 FTDDGGYMAHKTINYEAIETLLKARIKYYSGQAMRNQOV-GNSEIITSVRYGKGALKAT 695

QY 295 DTGTDTRTQGGIGVYVSNTPNLKLVNDKVLHMGRAHKNQOYRAAVLTTDGVINYTSD 354

DB 696 DTGDRTRTSGVAVIEGNNPNSDLKASDRVYVNMGAHKNQAYRPLLLTDDNGIKAYHSD 755

QY 355 QGAP--VAMTDENGDLYLSSHNLVVNGKEADTAQGYANPDVSGYLAVVWPGVAGSDQD 412

DB 756 QEAAGLVRYTNDRGELIFTA-----AD--INGYANPDVSGYLAVVWPGVAGSDQD 803

QY 413 ARTAPSTERNKSGNSAYRTNAFOSNVIFPAFSNVYTPTKESERANVRITAOADFPASIG 472

DB 804 VRVAASTPSTDGKSVHQNAALDSRVNFEGFSNFQAFATKKEEYTNVVIKKNVDFAEWG 863

QY 473 FTSFENAPQVNSKDKTELDSTDNGYAFDTRVDYDLGNSERNKYGTDEDLRNAIQALHKG 532

DB 864 VTFDEAPQVNSVSTDGSLDSVQNGYAFDTRVDYDLGNSERNKYGTADDLVKAIALHSG 923

QY 533 LQVMADWVPDQ 543

DB 924 IKVMADWVPDQ 934

RESULT 3

GTFD_STRMU

ID GTFD_STRMU STANDARD; PRT: 1462 AA.

AC P49331: O69383; O69386; O69392; O69398;

DT 01-FEB-1996 (Rel. 33, Created)

DT 15-JUL-1999 (Rel. 38, Last sequence update)

DT 15-JUL-1999 (Rel. 38, Last annotation update)

DE Glucosyltransferase-S precursor (EC 2.4.1.5) (GTF-S) (Dextranucrase)

Db 825 VPGASDQDVRVAASKANATQVYESSALOSQLYBGFNSFQDFTVKDSYTNKKIA 884
 QY 463 QNADFPASIGFTSFEMAPQYNSKDRFTFLDSIIDNGYATDRYOLGNSPKNYGTDEDLR 522
 Db 885 QNQLFKSGWTSFEMAPQYNSKDRFTFLDSIIDNGYATDRYOLGNSPKNYGTDEDLR 944
 QY 523 NAIQALHKAAGLOVADWVDPQ 543
 Db 945 NNVALLHKSGIQTADWVDPQ 965
 RESULT 4
 ID GTFL_STRDO STANDARD; PRT; 1592 AA.
 AC P27470:
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE Glucosyltransferase-I precursor (EC 2.4.1.5) (GTF-I) (Dextranucrase)
 DE (Sucrose 6-glucosyltransferase).
 OS Streptococcus downei (Streptococcus sobrinus).
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1317;
 RN 11
 RP SEQUENCE FROM N.A.
 RC STRAIN=6715;
 RX MEDLINE=91123227; PubMed=1704006;
 RA Abo H., Matsumura T., Kodama T., Ohta H., Fukui K., Kato K.,
 RT Kagawa H.;
 RT "Peptide sequences for sucrose splitting and glucan binding within
 Streptococcus sobrinus glucosyltransferase (water-insoluble glucan
 synthetase)".
 RL J. Bacteriol. 173:989-996(1991).
 CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT
 CC TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE
 CC OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIANE THE
 CC AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.
 CC -1- CATALYTIC ACTIVITY: Sucrose + [(1.6)-alpha-D-glucosyl](N) - D-
 CC fructose + [(1.6)-alpha-D-glucosyl](N+1).
 CC -1- SUBCELLULAR LOCATION: secreted.
 CC -1- DISEASE: DENTAL CARIES.
 CC -1- MISCELLANEOUS: GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA
 CC 1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES
 CC WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH
 CC FORMS OF GLUCANS.
 CC -1- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-
 CC BINDING PROTEIN FROM S.MUTANS.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).

 DR EMBL; D90213; BAA14241.1; -;
 DR PIR; A38175; A38175.
 DR HSSP; P00695; 2HEE.
 DR InterPro; IPR002479; CW_binding.
 DR InterPro; IPR003318; Glyco_hydro_70.
 DR Pfam; PF01473; CW_binding_1; 16.
 DR Pfam; PF02324; Glyco_hydro_70; 1.
 KW Transferase; Glycosyltransferase; Signal; Repeat; Dental caries.
 FT SIGNAL 1 38
 FT CHAIN 39 1592
 FT DOMAIN 39 1044
 FT CATALYTIC (APPROXIMATE).
 FT DOMAIN 1093 1592
 FT GLUCAN-BINDING (APPROXIMATE).
 FT DOMAIN 1093 1592
 FT REPEAT 1093 1142
 FT REPEAT 1158 1207
 FT REPEAT 1222 1272

FT REPEAT 1287 1337 4.
 FT REPEAT 1402 1451 5.
 FT REPEAT 1514 1563 6.
 FT REPEAT 1577 1592 7 (INCOMPLETE).
 SQ SEQUENCE 1592 AA; 176167 MW; BC0A66D079351BCF CRC64;
 Query Match 49.3%; Score 1406; DB 1; Length 1592;
 Best Local Similarity 52.6%; Pred. No. 1.1e-73;
 Matches 288; Conservative 83; Mismatches 153; Indels 24; Gaps 9;
 QY 5 GQEFLLANDIDNSNPVQAEQLNWLKYLNFQTTTANNDOANFDSVRVDAPDNIDADLMN 64
 Db 399 GYDELLANDVDSNPVQAEQLNWLKYLNFQTTTANNDOANFDSVRVDAPDNIDADLMN 458
 QY 65 IADQYENAAVGMDS-DAVSNKHINLELDANHADPEYFNKIGNPQLTMDTINKSNHGLS 123
 Db 459 ISSDYLAAGYDIDKNKNANNNHVSIVYEAWSDDPTFLHDDGNLMMNDKFRUSLWLSLA 518
 QY 124 DATN-RWGLDAIVHQSADRENNSNENNVIPNYSFYRAHDNNSQDQIONAIR-DVYTKDY 181
 Db 519 KPFDVRSGLNPLHNSLDREVDDREVEVPSPSPARAHDSVQDIIIRDIKAEINPNSF 578
 QY 182 -HTFTFEDEOKGIDAIYQDNSTVKKYKYNLYNPASVAILLTNKTIPRYVYGYDLYDQ 240
 Db 579 GYSFTQREIDQAFKIYNEIDLKSKDKYTHYNVPLSYTLITNKGSIPIRYVYGYDLYDQ 638
 QY 241 YMEHOFRYDYLNLKSRVYVAGQSMQYMSVGNNNILTSVRYGKGMATATDTGTDE 300
 Db 639 YMANKTVNYDAIESLKAARKYVAGQAMQYQI-NGEILLTSVRIGKALKQSDKGDAT 697
 QY 301 TRTQGGVYVSNTPNLKLGYNKVV-LHMGAAHKNQYRAAVLTFTDGVINTYSDQAGP- 958
 Db 698 TRTSGVGVWGMGNQENESL--DGKVALNMGAHANQEYRALMVSTKDGVTATYATDADASK 755
 QY 359 ---VMTDENGDIYLSHNLVYNGKEADTRVQGYANPDVSGYLAVWVPGASNDQDART 415
 Db 756 AGLVKRFDENGYLYFLNDDL-----KGVANPQVSGFLQVWVPGAAADDQDIRV 803
 QY 416 APSTEKNSGNSAVRTNAAPDSNVIFFAFSNFVTPTTKESEANVRIAQNAADFFASLGFTS 475
 Db 804 AASDASTDCKSLHQDAAMDSRVAFEGFSNFQSFANKEEYTVNVIANNVXKPVSWGITD 863
 QY 476 FEMAPQVSSKDRFTFLDSIIDNGYATDRYOLGNSPKNYGTDEDLNATQALHKAAGLOV 535
 Db 864 FEMAPQVSSDQGFQDSVIQNGYATDRYOLGNSKANKYGTADQLVKAIKALHANGLVK 923
 QY 536 MADWVDPQ 543
 Db 924 MADWVDPQ 931
 RESULT 5
 ID GTFL_STRDO STANDARD; PRT; 1597 AA.
 AC P11001;
 DT 01-JUL-1989 (Rel. 11, Created)
 DT 01-JUL-1989 (Rel. 11, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE Glucosyltransferase-I precursor (EC 2.4.1.5) (GTF-I) (Dextranucrase)
 DE (Sucrose 6-glucosyltransferase).
 GN GTFI.
 OS Streptococcus downei (Streptococcus sobrinus).
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1317;
 RN 11
 RP SEQUENCE FROM N.A.
 RC STRAIN=MFE28;
 RX MEDLINE=87308014; PubMed=3040686;
 RA Ferretti J.J., Gilpin M.L., Russell R.R.B.;
 RT "Nucleotide sequence of a glucosyltransferase gene from Streptococcus
 sobrinus MFE28.";

J. Bacteriol. 169:4271-4278(1987).
 CC -!- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT
 CC TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE
 CC OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE
 CC AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.
 CC -!- CATALYTIC ACTIVITY: Sucrose + {(1,6)-alpha-D-glucosyl}(N) - D-
 CC fructose + {(1,6)-alpha-D-glucosyl}(N+1).
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- DISEASE: DENTAL CARIES.
 CC -!- MISCELLANEOUS: GTF-1 SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA
 CC 1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES
 CC WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH
 CC FORMS OF GLUCANS.
 CC -!- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-
 CC BINDING PROTEIN FROM S.MUTANS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M17391; AAC53063.1; -
 CC InterPro: IPR002479; CW_binding.
 CC InterPro: IPR003318; Glyco_hydro_70.
 CC Pfam: PF01473; CW_binding_1; 19.
 CC Pfam: PF02324; Glyco_hydro_70; 1.
 CC Transferase; Glycosyltransferase; Signal; Repeat; Dental caries.
 CC SIGNAL 1 38 POTENTIAL
 CC CHAIN 39 1597 GLUCOSYLTRANSFERASE-I.
 CC DOMAIN 39 1050 CATALYTIC (APPROXIMATE).
 CC FT DOMAIN 1099 1597 GLUCAN-BINDING (APPROXIMATE).
 CC FT DOMAIN 1099 1597 1.25 A, 2 B, AND 5 AC REPEATS.
 CC FT REPEAT 1099 1597 A REPEAT.
 CC FT REPEAT 1163 1213 AC REPEAT.
 CC FT REPEAT 1227 1272 AC REPEAT.
 CC FT REPEAT 1292 1342 AC REPEAT.
 CC FT REPEAT 1352 1399 B REPEAT.
 CC FT REPEAT 1406 1455 AC REPEAT.
 CC FT REPEAT 1465 1512 B REPEAT.
 CC FT REPEAT 1519 1568 B REPEAT.
 CC FT REPEAT 1582 1597 A REPEAT (INCOMPLETE).
 CC SEQUENCE 1597 AA; 177080 MW; B9E86A20068798E CRC64;
 Query Match 49.3%; Score 1405; DB 1; Length 1597;
 Best Local Similarity 52.6%; pred. No. 1.3e-73;
 Matches 288; Conservative 82; Mismatches 154; Indels 24; Gaps 9;
 QY 5 GQEFFLANDNSNPVQAOBOLNMLYLLNFGTTTANNQDANFDSVRVDADNDADLMN 64
 DB 405 GYELLANDVNSNPVQAOBOLNMLYLLNFGTTTANNQDANFDSVRVDADNDADLLQ 464
 QY 65 IAQDFNAYGMD-SDAVSNKHINLEWADPEYFNKIGNPOLTTDDTKNSLNHGLS 123
 DB 465 ISSDYLAAGYIDKNNKANNHVSIVEAWSNDPTYLHDDGDNLMNDKPFSLMSLSA 524
 QY 124 DATN-RWGIDALVHQSADRENNSNVVPIPNYSFVRAHDNNSODQIGNAIR-DVTGKDY 181
 DB 525 RPLDKRSGNLPHNSLDREVDDREVEVTSYSEFARADHSEVDQLIRDKAEINPNAF 584
 QY 182 -HTTFEBEQGIDAYIQDQNSTVVKYNNLYNPASVAILLTKNTPIPVYGDLYDGGQ 240
 DB 585 GYSFQDEIDQAFKIYNEDLTKDKYTHNVPISYLLTNKSGIPRVYGDMEFTDDGQ 644
 QY 241 YNEHOTRYDTLTNLKSRVKYVAGQSQWMSVGGNNILTSVRYGKAGMTATDTGTD 300
 DB 645 YNANKTYNDAIESLTKARKMYVAGQSQWYQYI-NGEILTSVRYGKALKQSDKQAT 703
 QY 301 TETQGIYVVSNTPLKLGVDNKVY-LANGAAHKKQOYRAAVLTDTGVDVINYTSQDQAP- 358
 DB 358

DB 704 TETSGVVMGQPNFSL--DGKVVVALNMGAAHQEYRALMVSTKCGVATYATDADASK 761
 QY 359 ---VAMTENGDLYLSSNHLVYNGKEADTAVQGVANPDVSGYLAVVWVPGASNDQDART 415
 DB 762 AGLVARTDENGYLELNDL-----KGVANPQVSGFLQVWVPGAADQDQDIRV 809
 QY 416 APSTKNSGNSAYRTNAAFDSNVIFEAFSNFYVTPTKESERANVRVIAQNAOFFASLGFTS 475
 DB 810 AASDFASTDGKSLHODAAADSRYWFEFGSNFQSFATKBEETVNVVIANNVKPFVSWGITD 869
 QY 476 FEMAPQYSSKDKRTFLDSTIDNGYAFTRIDYLGMSPEPKYGTDEDLRMAIQALHKAGLOV 535
 DB 870 FEMAPQYSSSTDQGLDSVIONGYAFTRIDYLGMSKANKYGTADQLVKRAIKALHAKGLKV 929
 QY 536 MADWVPDQ 543
 DB 930 MADWVPDQ 937
 RESULT 6
 GTF-S_STRDO STANDARD; PRT: 1365 AA.
 ID GTF-S_STRDO STANDARD; PRT: 1365 AA.
 AC F29336;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DT 01-APR-1993 (Rel. 25, Last annotation update)
 DE Glucosyltransferase-S precursor (EC 2.4.1.5) (GTF-S) (Dextranucrase)
 DE (Sucrose 6-glucosyltransferase).
 GN GTF-S.
 OS Streptococcus downei (Streptococcus sobrinus).
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1317;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN-MFE28;
 RC MEDLINE=90316665; PubMed=2142479;
 RA Gilmore K.S., Russell R.R., Ferretti J.J.
 RT Analysis of the Streptococcus downei gtfS gene, which specifies a
 RT glucosyltransferase that synthesizes soluble glucans.;
 RL Infect. Immun. 58:2452-2458(1990).
 CC -!- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT
 CC TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE
 CC OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE
 CC AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.
 CC -!- CATALYTIC ACTIVITY: Sucrose + {(1,6)-alpha-D-glucosyl}(N) - D-
 CC fructose + {(1,6)-alpha-D-glucosyl}(N+1).
 CC -!- ENZYME REGULATION: GLUCAN SYNTHESIS BY GTF-S IS INDEPENDENT OF
 CC PRIMER GLUCAN UNLIKE GTF-1.
 CC -!- DISEASE: DENTAL CARIES.
 CC -!- MISCELLANEOUS: GTF-S SYNTHESIZES WATER-SOLUBLE GLUCANS (ALPHA
 CC 1,6-GLUCOSE).
 CC -!- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-
 CC BINDING PROTEIN FROM S.MUTANS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M30943; AAA26898.1; -
 CC PIR: A41483; A41483.
 CC InterPro: IPR002479; CW_binding.
 CC InterPro: IPR003318; Glyco_hydro_70.
 CC Pfam: PF01473; CW_binding_1; 10.
 CC Pfam: PF02324; Glyco_hydro_70; 1.
 CC Transferase; Glycosyltransferase; Signal; Repeat; Dental caries.
 CC SIGNAL 1 36 OR 37 (POTENTIAL).
 CC CHAIN 37 1365 GLUCOSYLTRANSFERASE-S.
 CC DOMAIN 37 1050 CATALYTIC (APPROXIMATE).
 CC

```
FT DOMAIN 1083 1365 GLUCAN-BINDING (APPROXIMATE).
FT DOMAIN 1083 1365 4.5 X TANDEM REPEATS.
FT REPEAT 1083 1131 1.
FT REPEAT 1150 1199 2.
FT REPEAT 1225 1274 3.
FT REPEAT 1289 1339 4.
FT REPEAT 1353 1365 5 (INCOMPLETE).
SQ SEQUENCE 1365 AA; 151590 MW; 167296B5A2B8C476 CRC64;

Query Match 42.8%; Score 1219.5; DB 1; Length 1365;
Best Local Similarity 46.8%; Pred. No. 5,4e-63;
Matches 259; Conservative 82; Mismatches 181; Indels 31; Gaps 9;

QY 3 YRGEFLAIDINDSNPVVQAEQLNLYLAFNGTITANNQANFDSVRVADPNIDADL 62
DB 387 YAGYELLANDVDSNPVQAEQLNLYLAFNGTITANNQANFDSVRVADPNIDADL 446
QY 63 MNTAQDYFNAAYGMD-SDAVSNKHINLEHDWNADEYFNKIGNPQLTMDDTIKNSLHG 121
DB 447 LQIQDYKAKYGTQDQNEKNAIDHLSILEAWGNDNDYVYKQDNFSLSDNDQSGMLKA 506
QY 122 LSDAT-NRGLDAIVHQSLADRENSTENVIPNYSFVRAHDNNSQDQTONAIRDVTGR- 179
DB 507 FGVASARGNLSNATAGLNKRSANPDSDPV-PNYVFIKRDSEYQTRAKIIRKLGRT 565
QY 180 ---DYHTFFEDQKIDAYIQDNSTVKKYMLNYPASYAILLNKQIPRYVYGDLYT 236
DB 566 NADCLTNLTLDLNAKAFDIYNQDMNATKVVYPNNLPNAYAWMLQNKDVTVRYVYGDLYT 625
QY 237 DGGQYHEHTRYDTLTNLLSKSVKVVAGQSM---QTSVSGVGNNNILTSVRVGKAMTA 293
DB 626 DNGQYMATKTPFFYNAETLKGHIKVVAGQAVSYVQDWSG-----ILTSVRIGKANS 681
QY 294 TQGTGTETQGVVVSNTPLKGVNDKVLHMGAAHKNQYRAAVLTMTTGTGVINYTS 353
DB 682 SDAGNTEINSGHALLINRNPNFRAYN-LTLNMGAAHKSQAYRPLLSTKDGATYLN 739
QY 354 D---QGAPVAMTDENGDIYLSHNLVNGKEADVAQVANDPVDVSGYLAVVVPVQASDN 410
DB 740 DSDVDSDQYKTYDSQNLFSASEL-----QSVANAQVSGMIQYVVPVGAADN 787
QY 411 QDARTAPSTEKNSGSAYRTNAAFDSNVIFAEFSNFTPTRESERANVRIAGNADFEAS 470
DB 788 QDVRTSPQATKQDNGIYHQSDALDSQVIEGFSQAFQSPQDQVNAVIAKNGDLFKS 847
QY 471 LGFTSFEMAPQYSSKDRFTLDSITDNGVAFYDRLGMSPEPNKYGTDEDLRNAIQALHK 530
DB 848 WGITQFEMAPQYSSYSDGTFDLSVILNGVAFSDRYDLAMSKNKNYKSGKDLANAIGLOS 907
QY 531 AGIQVADWVDPQ 543
DB 908 AGIKVLSDELVPNQ 920

RESULT 7
APU_THETU STANDARD; PRT; 1861 AA.
AC F38536;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Amylopullulanase precursor (Alpha-amylase/pullulanase) (Pullulanase type II) [includes: Alpha-amylase (EC 3.2.1.1) (1,4-alpha-D-glucan glucanohydrolase); Pullulanase (EC 3.2.1.41) (1,4-alpha-D-glucan glucanohydrolase) (Alpha-dextrin endo-1,6-alpha-glucosidase)].
GN AMVB.
OS Thermoanaerobacter thermosulfurogenes (Clostridium thermosulfurogenes).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Thermoanaerobacter group; Thermoanaerobacterium.
OX NCBI_TaxID=33950;
RN [1]
```

```
RP SEQUENCE FROM N.A.
RX STRAIN-DSM 3896 / EMBL;
RA MEDLINE=94252998; PubMed=8195085;
RA Matuschek M., Burchhardt G., Sahm K., Bahl H.;
RT "Pullulanase of Thermoanaerobacterium thermosulfurigenes EMI (Clostridium thermosulfurigenes): molecular analysis of the gene, composite structure of the enzyme, and a common model for its attachment to the cell surface";
RT J. Bacteriol. 176:3295-3302(1994).
RL -I- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-alpha-glucosidic linkages in oligosaccharides and polysaccharides.
CC -I- CATALYTIC ACTIVITY: Hydrolysis of (1->6)-alpha-D-glucosidic linkages in pullulan and in amylopectin and glycogen, and the alpha- and beta-limit dextrins of amylopectin and glycogen.
CC -I- SUBCELLULAR LOCATION: CELL-BOUND. IT C-TERMINUS MAY SERVE AS AN S-LAYER ANCHOR.
CC -I- PTM: GLYCOSYLATED.
CC -I- SIMILARITY: BELONGS TO FAMILY 13 OF GLYCOSYL HYDROLASES, ALSO KNOWN AS THE ALPHA-AMYLASE FAMILY.
CC -I- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -I- SIMILARITY: CONTAINS 3 S-LAYER HOMOLGY (SLH) DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M57692; AAB00841.1; -.
DR HSP; Q08751; LBV2.
DR InterPro; IPR000461; Alpha-amylase.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR001119; SLH.
DR InterPro; IPR004185; alpha-amylase_N.
DR InterPro; IPR004193; isoamylase_N.
DR Pfam; PF00128; alpha-amylase; 1.
DR Pfam; PF02806; alpha-amylase_C; 1.
DR Pfam; PF02903; alpha-amylase_N; 1.
DR Pfam; PF00041; fn3; 2.
DR Pfam; PF02922; isoamylase_N; 1.
DR Pfam; PF00395; SLH; 3.
DR SMART; SM00060; FN3; 1.
DR PROSITE; PS01072; SLH_DOMAIN; 3.
KW Hydrolase; Glycosidase; Carbohydrate metabolism; Signal; Repeat; Multifunctional enzyme; Glycoprotein.
KW SIGNAL 1 35
FT CHAIN 36 1861 AMYLOPULLULANASE.
FT DOMAIN 928 1018 FIBRONECTIN TYPE-III 1.
FT DOMAIN 1157 1248 FIBRONECTIN TYPE-III 2.
FT ACT_SITE 628 628 BY SIMILARITY.
FT ACT_SITE 657 657 BY SIMILARITY.
FT ACT_SITE 734 734 BY SIMILARITY.
FT DOMAIN 1681 1739 SLH 1.
FT DOMAIN 1740 1803 SLH 2.
FT DOMAIN 1804 1861 SLH 3.
FT CONFLICT 1734 1734 D->E (IN REF. 1: AAB00841).
SQ SEQUENCE 1861 AA; 206104 MW; 06C23070E453B574 CRC64;

Query Match 5.5%; Score 156.5; DB 1; Length 1861;
Best Local Similarity 19.28; Pred. No. 0.19;
Matches 126; Conservative 94; Mismatches 197; Indels 239; Gaps 35;

QY 12 NDIDNSNPVQAEQLNLYLAFNGTITANNQANFDSVRV-----AFDNIDA 60
DB 1212 NYIDTS-VINGVTYNYKVVAVD---LSFNRTESNVVTIKPDVPIKVFNVTPD-YTP 1265
QY 61 DLANIADQDFNAAYGMDSDAVSNKHINLE-----DWNHAD-PEYFN 101
DB 1266 DAVNLASTFPNATWDFPSAQOMTKIDNTYISITLDSGTQIEYKARGSWDKVEKDEYN 1325
```

QY 102 KI-----GNPQITMDOTIKNSLHGLSDATNRWGLDAIVHQSLADRENNSTENV 150
 DB 1326 EFASNRKVTIVNGNNEMTINDTV-----YRW-----RD----- 1354
 QY 151 VIPNTSEVRAHDNNSQDQIQNAIRDVTGKDYHTTFEDEQKGI-----DAYIQDQNSTV 204
 DB 1355 -IP--IFYIYSPSNTVDSNISTMEVKNTY-----KGAKVTINGDSFVQDKNGVF 1402
 QY 205 KRYNLYNPASAYAILLTNKDTI---PRVYGDLYPDGGYMEHQRYVDTLTNLKSRVK 261
 DB 1403 TK----DVSNTGV---NKIKIHVEP-----NDGSGVYGNDOGR---ITELTKD-IE 1442
 QY 262 YVAGGSMQTHSVGGNNILTSVRVYKGAFTDGTDETRQGGIGVVVSNTPNLK--LG 319
 DB 1443 IDVIRQNNSSGSGTGNNTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 1496
 QY 320 V-----NKKVVLHMGAAHKNQYRAAVLTITDGVINYTSD 354
 DB 1497 VITKNGNVTITLDAGKAKDLIVNSKDKKVVFDITIGSGQQ---KVQIISKDILDTSA 1553
 QY 355 QGAPVAMTDENGDLKLS-----SHNLVYNGKEADTAVQYANPDVSGYLVNVPV----- 405
 DB 1554 NGKQIVIKSDNASIALTKDALNQNTQNG---VNVSIKDKNGKPNVTNYVLSNVVDITIS 1610
 QY 406 GASDNDQARTATPSTKNSGNSAYRTNAAFDSNVIFEAFSNFYVTPTKES-----ERA 457
 DB 1611 GISGNTVIAKPEVTILN-----ISKANDPKV-----AVYYNFTTNQWYVGGKVDAS 1659
 QY 458 NVRIQADOFFASLGTSFEMAPQYNSKD-----RTFLDSTIDNGYAFTRDY 505
 DB 1660 SGITITNATHFSQ---YAAFEYDKFTDKDNWAKDVIEVLSRSHIVEGMDTQY----- 1711
 QY 506 DLGNSPKNYGT-----DE-----DLR-----NAIQALHKAGL 533
 DB 1712 -----BPNKTVTRAETAMILRLINIKDEYSGSPDKSGDYANATBAAYKAGI 1762

RESULT 8
 AMY_BACLI STANDARD; PROT; 512 AA.
 AC P06278;
 DT 01-JAN-1988 (Rel. 06, Created)
 DT 01-JAN-1988 (Rel. 06, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Alpha-amylase precursor (EC 3.2.1.1) (1,4-alpha-D-glucan
 DE glucanohydrolase).
 GN AMYS OR AMYL.
 OS Bacillus licheniformis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=1402;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 27811;
 RX MEDLINE=86111694; PubMed=2418011;
 RA Yuuki T., Nomura T., Tezuka H., Tsuboi A., Yamagata H.,
 RA Tsukagoshi N., Ueda S.;
 RT "Complete nucleotide sequence of a gene coding for heat- and
 RT pH-stable alpha-amylase of Bacillus licheniformis: comparison of the
 RT amino acid sequences of three bacterial liquefying alpha-amylases
 RT deduced from the DNA sequences";
 RL J. Biochem. 98:1147-1156(1985).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=86195857; PubMed=3009417;
 RA Gray G.L., Mainzer S.E., Rey M.W., Lamsa M.H., Kindle K.L.,
 RA Carmona C., Reuquadt C.;
 RT "Structural genes encoding the thermophilic alpha-amylases of
 RT Bacillus stearothermophilus and Bacillus licheniformis";
 RL J. Bacteriol. 166:635-643(1986).
 RN [3]
 RP SEQUENCE OF 1-104 FROM N.A.
 RX MEDLINE=84185455; PubMed=6609154;

RA Stephens M.A., Ortlepp S.A., Ollington J.F., McConnell D.J.;
 RT "Nucleotide sequence of the 5' region of the Bacillus licheniformis
 RT alpha-amylase gene: comparison with the B. amyloliquefaciens gene.";
 RL J. Bacteriol. 158:369-372(1984).
 RN [4]
 RP SEQUENCE OF 1-29 FROM N.A.
 RX MEDLINE=89213924; PubMed=2540150;
 RA Laiole B.M., Chambliss G.H., McConnell D.J.;
 RT "Bacillus licheniformis alpha-amylase gene, amyL, is subject to
 RT promoter-independent catabolite repression in Bacillus subtilis.";
 RL J. Bacteriol. 171:2435-2442(1989).
 RN [5]
 RP SEQUENCE OF 30-47.
 RX MEDLINE=82098050; PubMed=6172418;
 RA Kuhn H., Fietzek P.P., Lampen J.O.;
 RT "N-terminal amino acid sequence of Bacillus licheniformis
 RT alpha-amylase: comparison with Bacillus amyloliquefaciens and
 RT Bacillus subtilis Enzymes.";
 RL J. Bacteriol. 149:372-373(1982).
 RN [6]
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).
 RC STRAIN=ATCC 27811;
 RX MEDLINE=95182462; PubMed=7877175;
 RA Machius M., Wiegand G., Huber R.;
 RT "Crystal structure of calcium-depleted Bacillus licheniformis alpha-
 RT amylase at 2.2-A resolution.";
 RL J. Mol. Biol. 246:545-559(1995).
 CC -I- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-alpha-glucosidic
 CC linkages in oligosaccharides and polysaccharides.
 CC -I- BIOTECHNOLOGY: Used in the food industry for high temperature
 CC liquefaction of starch-containing mash and in the detergent
 CC industry to remove starch. Sold under the name Termamyl by
 CC Novozymes.
 CC -I- MISCELLANEOUS: ABLE TO WORK AT RELATIVELY HIGH (ALKALINE) PH
 CC VALUES (UP TO PH 11) AND AT HIGH TEMPERATURES (UP TO 100 DEGREE
 CC CELSIUS).
 CC -I- SIMILARITY: BELONGS TO FAMILY 13 OF GLYCOSYL HYDROLASES, ALSO
 CC KNOWN AS THE ALPHA-AMYLASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X03336; CAA26981.1; -;
 DR EMBL; M38570; AAA22226.1; -;
 DR EMBL; M13256; AAA22240.1; -;
 DR EMBL; K01984; AAA22193.1; -;
 DR EMBL; M26412; AAA22237.1; -;
 DR EMBL; A17930; CAA01355.1; -;
 DR PIR; A00844; ALBSL.
 DR PIR; B24549; B24549.
 DR PIR; A26151; A26151.
 DR PDB; 1BPL; 17-AUG-96.
 DR PDB; 1VJS; 12-MAR-97.
 DR InterPro; IPR000461; Alpha_amylase.
 DR Pfam; PF00128; alpha-amylase; 1.
 DR Hydrolase; Glycosidase; Carbohydrate metabolism; Signal; 3D-structure.
 KW SIGNAL 1 29
 FT CHAIN 30 512
 FT ACT_SITE 260 260
 FT ACT_SITE 264 264
 FT ACT_SITE 357 357
 FT CONFLICT 38 38 Q -> Y (IN REF. 5).
 FT CONFLICT 163 163 R -> L (IN REF. 2).
 FT CONFLICT 339 339 S -> G (IN REF. 2).
 FT CONFLICT 349 349 A -> S (IN REF. 2).
 SQ SEQUENCE 512 AA; 58549 MW; D8BB77759CD4C482 CRC64;

Query Match 5.5%; Score 155.5; DB 1; Length 512;
Best Local Similarity 23.3%; Pred. No. 0.039;
Matches 79; Conservative 42; Mismatches 117; Indels 101; Gaps 15;

QY 7 EFLANDINSPVQAEQLNWLIIYLLNFGTTTANNQANFDSVRDAPDNIDADLMNIA 66
DB 223 DYLMYADVDYDHPDVAEAKRW-----GTWYAN--ELQDGFRLDAV----- 262
QY 67 QDYFNAAYGMSDAVSNKHI--NILEDW-NHADPEYFNKIGNPOLMTDDTIKNSLNHGLS 123
DB 263 -----XKIKSFRLDWNHVR-----EKTGKEMFTVAEYQNDLGL-ALE 300
QY 124 DATNRGLDAIVHQSLADRENNSTENVIPNVSFVRAHND-----NSQDQIQNAI 173
DB 301 NYLKNTFN-----NHSVFDVPL-HYOFHAASTQGGYDMRKILNSTVVSXKHL 347
QY 174 RDVTGKDYHTFEDBQKIDAYIQDNSTVKKYINLPASVAILTNKDTIPRVYGD 233
DB 348 KATVFDNH-----DTQPG-----QSLESTVQTFWK---PLAYAFILTRSGYPOVFGD 394
QY 234 LYPDGGYMEHOFYVDTLTNLKSRVYVAGGQ-----SMQTMVSGVGNLIL 281
DB 395 MYCTGDSQREIPALXKHKIEPIKARKQAYGAQHDFDHDHIVGWTREGDSVANSGLA 454
QY 282 TSVRYCKGA-----WATDGTDETRTQIGVYVSN 312
DB 455 ALITDGGGAKRMVGRQAGETWHDITGNRSEPVVINS 493

RESULT 9
ID ANY_BACAM STANDARD; PRT; 514 AA.
AC P00692;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Alpha-amylase precursor (EC 3.2.1.1) (1,4-alpha-D-glucan
glucanohydrolase).
OS Bacillus amyloliquefaciens.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1390;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JH;
RX MEDLINE=8310808; PubMed=6185474;
RA Takkinen K., Pettersson R.F., Kalkkinen N., Palva I., Soederlund H.,
RA Kaerlaeinen L.;
RT "Amino acid sequence of alpha-amylase from Bacillus amyloliquefaciens
deduced from the nucleotide sequence of the cloned gene.";
RL J. Biol. Chem. 258:1007-1013(1983).
RN [2]
RP SEQUENCE OF 32-222.
RX MEDLINE=80241725; PubMed=6156671;
RA Chung H.S., Friedberg F.;
RT "Sequence of the N-terminal half of Bacillus amyloliquefaciens alpha-
amylase.";
RL Biochem. J. 185:387-395(1980).
RN [3]
RP SEQUENCE OF 1-96 FROM N.A.
RX MEDLINE=82051296; PubMed=6170539;
RA Palva I., Pettersson R.F., Kalkkinen N., Lehtovaara P., Sarvas M.,
RA Soederlund H., Takkinen K., Kaerlaeinen L.;
RT "Nucleotide sequence of the promoter and NH2-terminal signal peptide
region of the alpha-amylase gene from Bacillus amyloliquefaciens.";
RL Gene 15:43-51(1981).
RN [4]
RP SEQUENCE OF 1-39 FROM N.A.
RX MEDLINE=88137952; PubMed=2830166;
RA Ruohonen L., Hackman P., Lehtovaara P., Knowles J.K.C., Karaenen S.;
RT "Efficient secretion of Bacillus amyloliquefaciens alpha-amylase by
its own signal peptide from Saccharomyces cerevisiae host cells.";
RL Gene 59:161-170(1987).

CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-alpha-glucosidic
CC linkages in oligosaccharides and polysaccharides.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO FAMILY 13 OF GLYCOSYL HYDROLASES, ALSO
CC KNOWN AS THE ALPHA-AMYLASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: J01542; AAA22191.1; -
CC EMBL: V00092; CAA23430.1; -
CC EMBL: A20154; CAA01489.1; -
CC EMBL: M18424; AAA22192.1; -
CC PIR: A00843; ALBSN.
CC HSSP: P06278; 1VJS.
CC InterPro: IPR000461; Alpha_amylase.
CC Pfam: PF00128; alpha-amylase; 1.
KW Hydrolyase; Glycosidase; Carbohydrate metabolism; Signal.
FT SIGNAL 1 31
FT CHAIN 32 514 ALPHA-AMYLASE.
FT ACT_SITE 262 262 BY SIMILARITY.
FT ACT_SITE 266 266 BY SIMILARITY.
FT ACT_SITE 359 359 BY SIMILARITY.
FT CONFLICT 54 54 L -> I (IN REF. 2).
FT CONFLICT 64 64 I -> L (IN REF. 2).
FT CONFLICT 79 79 S -> D (IN REF. 2).
FT CONFLICT 84 84 G -> S (IN REF. 2).
SQ SEQUENCE 514 AA; 58403 MW; 3DF66B3FB5CCDE7E CRC64;

Query Match 5.2%; Score 149.5; DB 1; Length 514;
Best Local Similarity 22.3%; Pred. No. 0.087;
Matches 60; Conservative 35; Mismatches 103; Indels 71; Gaps 10;

QY 7 EFLANDINSPVQAEQLNWLIIYLLNFGTTTANNQANFDSVRDAPDNIDADLMNIA 66
DB 225 DYLMYADVDYDHPDVAETKKMGWYAN-----ELSLDGFRIIDAAKHK---FSEL 272
QY 67 QDYFNAAYGMSDAVSNKHNILEDW-NHAD--PEYFNKIGNPOLMTDDTIKNSLNHGLS 123
DB 273 RDWQAV-----RQATGKEMFTVAEYQNDLGLN-----KTSFNQSVF 315
QY 124 DATNRGLDAIVHQSLADRENNSTENVIPNVSFVRAHNDNSQDQIQNAIRDTVG 178
DB 316 DVELHENLQAASQGGYDMRLLDGTVVSRRHPEKAVTFVFNHDT----- 360
QY 179 KDYHTFTFEDBQKIDAYIQDNSTVKKYINLPASVAILTNKDTIPRVYGDLYTDG 238
DB 361 -----QPG-----QSLESTVQTFWK---PLAYAFILTRSGYPOVFGDYGMVTK 401
QY 239 GOYMEHQRTYDITLTNLKSRVYVAGGQ 267
DB 402 GTSPEKPEISLKDNIPIKARKEYAVGPQ 430

RESULT 10
ID ANY_BACAM STANDARD; PRT; 518 AA.
AC P19571;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Glucan 1,4-alpha-maltohexaosidase precursor (EC 3.2.1.98) (G6-amylase)
DE (Maltotetraose-producing amylase) (Exo-maltohexaohydrolase).
OS Bacillus sp. (strain 707).
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1416;

```

RN  [1]
RP  SEQUENCE FROM N.A., AND SEQUENCE OF 34-36.
RX  MEDLINE-88162814; PubMed-3258152;
RA  Tsukamoto A., Kimura K., Ishii Y., Takano T., Yamane K.;
RT  "Nucleotide sequence of the maltotetraose-producing amylose gene from
RT  an alkalophilic Bacillus sp. #707 and structural similarity to
RL  liquefying type alpha-amyloses."
RL  Biochem. Biophys. Res. Commun. 151:25-31(1988).
CC  -1- CATALYTIC ACTIVITY: Hydrolysis of 1,4-alpha-D-glucosidic linkages
CC  in amylose polysaccharides so as to remove successive
CC  maltotetraose residues from the non-reducing chain ends.
CC  -1- PATHWAY: DEGRADATION OF STARCH.
CC  -1- SUBCELLULAR LOCATION: Secreted.
CC  -1- SIMILARITY: BELONGS TO FAMILY 13 OF GLYCOSYL HYDROLASES, ALSO
CC  KNOWN AS THE ALPHA-AMYLASE FAMILY.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; M18862; AAA22231.1; -
DR  PIR; A27705; A27705.
DR  RSP; P06278; IVJS.
DR  InterPro; IPR000461; Alpha_amyase.
DR  Pfam; PF00128; alpha-amyase; 1.
KW  Hydrolase; Glycosidase; Carbohydrate metabolism; Signal.
FT  SIGNAL 1 33
FT CHAIN 34 518 GLUCAN 1,4-ALPHA-MALTOHEXAOSIDASE.
FT SEQUENCE 518 AA; 59009 MW; 3A961E21612682C4 CRC64;
SQ  -----
Query Match 5.1%; Score 144.5; DB 1; Length 518;
Best Local Similarity 20.7%; Pred. No. 0.17;
Matches 74; Conservative 56; Mismatches 113; Indels 115; Gaps 16;
QY 7 EFLANDIDNSFPVQAEQNW-LYLLNFGTITANNQANFDSVRVDPNDTADLMI 65
DQ 232 DYLMAADIDMDHPVEVNELRNGWVWTTTLG-----LDGFRIDAVKHK---YSF 278
QY 66 AQDYFN---AAYGMSDAVSNKHNILEDNADPEFNKIGNPQLTMDTTIKNSLHGL 122
DQ 279 TRDINIVRSATGKNMFAVA-----EFNK-----NDLGALENTLQ---KTNMHSV 321
QY 123 SDATNRWGLDAIVHQSIL--ADRENNSTENVIPNYSFVRADHNNSDQIQNAIRDVTGKD 180
DQ 322 -----FDVPLHLYNASKSGGYDMRNIFNGTVVQRHPS----- 356
QY 181 YHTTFDE-----QKGDYIQDQNSTVKYKYNLYNPASYAILLTKOYIPRYIGDLY 235
DQ 357 -HAVTFVDNHSQPEALESFVEE-----WEKPLAYALTLTREQYPSVFYGYDY 405
QY 236 ---TDGGYMEHPYRYDTLTNLKSRVKYVAG-----GOSQMTMSVGGNNNI 280
DQ 406 GIPTHGVPMKSK-----IDPILAEKQKAYGKQNDYLDHNNIIGTRGNTAHPMSGL 459
QY 281 LTVRYGKGA-----MTATDTGTDETRTQGGIGVVVSNTPNKLGLVN 321
DQ 460 ATIMSDGAGGKWNFVGNKAGQVSDITCNRTGTVTINADGWNFVSGSVSIWYN 517
RESULT 11
IDGT_BACCI
ID CDGT_BACCI STANDARD; PRT; 718 AA.
AC P30920;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Cyclomaltodextrin glucanotransferase precursor (EC 2.4.1.19)
DE (Cyclodextrin-glycosyltransferase) (CGTase).

```

```

OS  Bacillus circulans.
OC  Bacteria; Firmicutes; Bacillus/Clostridium group;
OC  Bacillus/Staphylococcus group; Bacillus.
OX  NCBI_TaxID-1397;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN-8;
RX  MEDLINE-91103970; PubMed-1368573;
RA  Nitschke L., Heeger K., Bender H., Schulz G.E.;
RT  "Molecular cloning, nucleotide sequence and expression in Escherichia
RT  coli of the beta-cyclodextrin glycosyltransferase gene from Bacillus
RT  circulans strain no. 8."
RL  Appl. Microbiol. Biotechnol. 33:542-546(1990).
RN  [2]
RP  X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
RC  STRAIN-8;
RX  MEDLINE-91171298; PubMed-1826034;
RA  Klein C., Schulz G.E.;
RT  "Structure of cyclodextrin glycosyltransferase refined at 2.0-A
RT  resolution."
RL  J. Mol. Biol. 217:737-750(1991).
RN  [3]
RP  X-RAY CRYSTALLOGRAPHY (3.4 ANGSTROMS).
RC  STRAIN-8;
RX  MEDLINE-90064533; PubMed-2531228;
RA  Hofmann B.E., Bender H., Schulz G.E.;
RT  "Three-dimensional structure of cyclodextrin glycosyltransferase from
RT  Bacillus circulans at 3.4-A resolution."
RL  J. Mol. Biol. 209:793-800(1989).
RN  [4]
RP  X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS).
RC  STRAIN-8;
RX  MEDLINE-98226626; PubMed-9558324;
RA  Schmidt A.K., Cottaz S., Briguez H., Schulz G.E.;
RT  "Structure of cyclodextrin glycosyltransferase complexed with a
RT  derivative of its main product beta-cyclodextrin."
RL  Biochemistry 37:5909-5915(1998).
RN  [5]
RP  X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS).
RC  STRAIN-8;
RX  MEDLINE-98409292; PubMed-9738912;
RA  Parsiegla G., Schmidt A.K., Schulz G.E.;
RT  "Substrate binding to a cyclodextrin glycosyltransferase and
RT  mutations increasing the gamma-cyclodextrin production."
RL  Eur. J. Biochem. 255:710-717(1998).
CC  -1- CATALYTIC ACTIVITY: Degrades starch to cyclodextrins by formation
CC  of a 1,4-alpha-D-glucosidic bond.
CC  -1- COFACTOR: BINDS TWO CALCIUM IONS.
CC  -1- SUBUNIT: MONOMER.
CC  -1- MISCELLANEOUS: CGTASE MAY CONSIST OF TWO PROTEIN DOMAINS: THE ONE
CC  IN THE AMINO-TERMINAL SIDE CLEAVES THE ALPHA-1,4-GLUCOSIDIC BOND
CC  IN STARCH, AND THE OTHER IN THE C-TERMINAL SIDE CATALYZES OTHER
CC  ACTIVITIES, INCLUDING THE RECONSTITUTION OF AN
CC  ALPHA-1,4-GLUCOSIDIC LINKAGE FOR CYCLIZING THE
CC  MALTOOLIGOSACCHARIDE PRODUCED.
CC  -1- SIMILARITY: BELONGS TO FAMILY 13 OF GLYCOSYL HYDROLASES, ALSO
CC  KNOWN AS THE ALPHA-AMYLASE FAMILY.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; X68326; CAA48401.1; -
DR  PIR; S23674; ALBSGC
DR  PDB; 1CGT; 31-JAN-94.
DR  PDB; 1CGU; 31-JAN-94.
DR  PDB; 3CGT; 27-MAY-98.
DR  PDB; 4CGT; 12-AUG-98.
DR  PDB; 5CGT; 12-AUG-98.

```

DR PDB; 6CGT; 14-OCT-98.
DR PDB; 7CGT; 12-AUG-98.
DR PDB; 8CGT; 14-OCT-98.
DR PDB; 9CGT; 14-OCT-98.
DR InterPro; IPR000461; Alpha_amylase.
DR InterPro; IPR002044; CBD_4.
DR InterPro; IPR002909; IPT_TIG.
DR Pfam; PF00128; alpha-amylase; 1.
DR Pfam; PF02806; alpha-amylase; 1.
DR Pfam; PF00686; CBD_4; 1.
DR Pfam; PF01833; TIG; 1.
DR ProDom; PD001568; CBD_4; 1.
KW Transferrase; Glycosyltransferase; Calcium; Signal; 3D-structure.
FT SIGNAL 1 34
FT CHAIN 35 718 CYCLOMALTODEXTRIN GLUCANOTRANSFERASE.
FT DOMAIN 35 172 A1.
FT DOMAIN 173 236 B.
FT DOMAIN 237 440 A2.
FT DOMAIN 441 528 C.
FT DOMAIN 529 614 D.
FT DOMAIN 615 718 E.
FT DISULFID 77 84
FT ACT_SITE 258 258
FT ACT_SITE 291 291
FT ACT_SITE 362 362
FT TURN 40 41
FT TURN 43 44
FT TURN 47 48
FT STRAND 51 53
FT HELIX 56 59
FT TURN 71 72
FT STRAND 73 73
FT TURN 75 76
FT TURN 80 81
FT STRAND 83 83
FT HELIX 88 96
FT TURN 97 98
FT HELIX 99 103
FT TURN 104 104
FT STRAND 107 109
FT STRAND 114 116
FT STRAND 121 123
FT TURN 124 125
FT STRAND 126 128
FT TURN 131 132
FT STRAND 136 142
FT TURN 144 146
FT HELIX 149 161
FT TURN 162 163
FT STRAND 165 170
FT TURN 172 173
FT STRAND 174 177
FT TURN 183 184
FT TURN 186 189
FT STRAND 191 193
FT TURN 194 195
FT STRAND 196 199
FT TURN 202 203
FT TURN 205 206
FT STRAND 209 210
FT STRAND 214 214
FT HELIX 220 225
FT TURN 227 228
FT STRAND 229 230
FT STRAND 231 234
FT TURN 236 237
FT HELIX 239 254
FT TURN 255 256
FT STRAND 259 262
FT HELIX 265 267
FT HELIX 270 281
FT TURN 282 283
FT STRAND 287 290

FT HELIX 301 309
FT STRAND 313 315
FT HELIX 317 327
FT TURN 328 329
FT HELIX 334 347
FT TURN 349 350
FT HELIX 351 353
FT STRAND 355 356
FT TURN 361 362
FT TURN 369 370
FT HELIX 373 385
FT STRAND 389 393
FT TURN 396 399
FT TURN 405 408
FT HELIX 420 428
FT TURN 429 430
FT HELIX 431 434
FT HELIX 436 440
FT STRAND 442 449
FT TURN 450 451
FT STRAND 452 459
FT TURN 460 461
FT STRAND 462 469
FT STRAND 476 478
FT STRAND 482 482
FT STRAND 488 490
FT TURN 493 498
FT STRAND 503 506
FT TURN 507 508
FT STRAND 509 510
FT STRAND 513 515
FT TURN 517 518
FT STRAND 520 525
FT STRAND 533 538
FT STRAND 541 542
FT TURN 544 545
FT STRAND 547 553
FT STRAND 561 564
FT TURN 565 566
FT STRAND 567 569
FT TURN 571 572
FT STRAND 574 578
FT TURN 579 580
FT STRAND 581 585
FT STRAND 592 600
FT TURN 601 602
FT STRAND 603 613

Query Match 5.0%; Score 141.5; DB 1; Length 718;

Best Local Similarity 21.8%; Pred. No. 0.39;

Matches 105; Conservative 55; Mismatches 184; Indels 137; Gaps 22;

QY 64 NIAQDYFNAAYGMDSDAVSN---KHINILEDWNHADPEYFNKIGNPQLTMDDTIKNSLHH 120

DB 203 NDTNGYFHHNGSGDSFSLGNGIYKNLYDLADFNHN-----ATIDKYFKDAIKL 251

QY 121 GLSDATNRGLDAIVHQSLADREN-----NSTENVVIPNYSF 157

DB 252 WLDWGYDGIKRVDAVKHPLGWQKSWMSIYAHKPVTFGEWFLGSAASADADNTDFANKSG 311

QY 158 VRAHDNNSQDIQONAIRDVTGDKYHTTFEDEQKGDIDAYIQDONSTVKKYNLYN----- 211

DB 312 MSLLDFRFSAVNRVFRDNTSNRY-----AUDSMI---NSTATDYNQVNDQVTFI 358

QY 212 -----IPASYAILLTNKDTPRVYTG-DLYTDGGQYMEHQTR----- 247

DB 359 DNHDMDFKTSAVNNRLEQALFTLTSGR-VPAIYVGTQELTGTGNGDPNRAKMPSESK 417

QY 248 -----YYDTLTNLLKSRVKNYAG-----CQSMQTSVGVGNMILTS 283

DB 418 STTAFNVISKLAPLRKSNPAIAYGSTQQRWINNDVYIERKFGKSVAVAV--NRNLSTS 475

QY 284 VRYGKGAMTATDGTDETRTGGIGVVVSNTPNLKLGYNVDKVLHMGAAHKNQOYRAVL 343

DB 476 ASI-TGLSTSLTGS-YTDVIG-GVLNNGNIITSTNGSINNF7LAAGAT-AWQYTTAETT 531
 QY 344 TTDGVINTSDQAPVAMTDENGDLVLSHNLVWVKEADTAOGYANPDVSGYLAVW 403
 DB 532 TTIGHVGPVMPKPGNVTTIDGRG--FGSTKGVYF6-----TTAVTGAA-----ITSW- 577
 QY 404 PVGASNDQARTAPSTERNKSGNSAYRINAFAFNSVIFAFSTPTKESRANVRIAQ 463
 DB 578 ----EDTIKIVIPSA--AGNVAVKVA--SGVNSNAYNFTIL--TGQVTVRFVV 625
 QY 464 N 464
 DB 626 N 626

RESULT 12

AMV2_ECOLI STANDARD; PRT; 495 AA.
 ID AMV2_ECOLI
 AC P26612; P78072;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Cytoplasmic alpha-amylase (EC 3.2.1.1) {1,4-alpha-D-glucan
 DE glucanohydrolase).
 GN AMYA OR B1927.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-JALL1;
 RX MEDLINE=9301517; PubMed=1400215;
 RA Raha M., Kawagishi I., Mueller V., Kihara M., Macnab R.M.;
 RT "Escherichia coli produces a cytoplasmic alpha-amylase, AnyA.";
 RL J. Bacteriol. 174:6644-6652(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12;
 RX MEDLINE=97251358; PubMed=9097040;
 RA Itoh T., Alba T., Baba T., Fujita K., Hayashi K., Inada T.,
 RA Isono K., Kasai H., Kimura S., Kitakawa M., Kitagawa M.,
 RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
 RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T.,
 RA Saito N., Sempel G., Seki Y., Sivasubramanian S., Tagami H.,
 RA Takeda J., Takemoto K., Wada C., Yamamoto Y., Horiuchi T.;
 RT "A 460-kb DNA sequence of the Escherichia coli K-12 genome
 corresponding to the 40.1-50.0 min region on the linkage map.";
 RL DNA Res. 3:379-392(1996).
 RN [4]
 RP SEQUENCE OF 1-5 FROM N.A.
 RC STRAIN-JALL1.
 RX MEDLINE=92407478; PubMed=1527488;
 RA Kawagishi I., Mueller V., Williams A.W., Irikura V.M., Macnab R.M.;
 RT "Subdivision of flagellar region III of the Escherichia coli and
 RT Salmonella typhimurium chromosomes and identification of two
 RT additional flagellar genes.";
 RL J. Gen. Microbiol. 138:1051-1065(1992).
 RN [5]
 RP SEQUENCE OF 475-495 FROM N.A.
 RC STRAIN-JALL1;
 RX MEDLINE=93381452; PubMed=8371104;

RA Raha M., Kihara M., Kawagishi I., Macnab R.M.;
 RT "Organization of the Escherichia coli and Salmonella typhimurium
 RT chromosomes between flagellar regions IIIa and IIIb, including a
 RT large non-coding region.";
 RL J. Gen. Microbiol. 139:1401-1407(1993).
 CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-alpha-glucosidic
 CC linkages in oligosaccharides and polysaccharides.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -!- SIMILARITY: BELONGS TO FAMILY 13 OF GLYCOSYL HYDROLASES, ALSO
 CC KNOWN AS THE ALPHA-AMYLASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; L01642; AAA23810.1; -.
 DR EMBL; AE000285; AAC74994.1; -.
 DR EMBL; D90833; BAA15755.1; -.
 DR EMBL; M85240; -; NOT ANNOTATED_CDS.
 DR EMBL; L13279; AAA82575.1; -.
 DR PIR; A45738; A45738.
 DR HSP; P06278; 1VJS.
 DR Ecogene; EGI1387; amyA.
 DR InterPro; IPR000461; Alpha_amylase.
 DR Pfam; PF00128; alpha-amylase; 1.
 KW Hydrolase; Glycosidase; Carbohydrate metabolism; Complete proteome.
 FT ACT_SITE 235 235
 FT ACT_SITE 239 239
 FT ACT_SITE 332 332
 FT CONFLICT 19 20 KL -> SS (IN REF. 1).
 FT CONFLICT 109 109 A -> V (IN REF. 1).
 FT CONFLICT 149 149 Q -> E (IN REF. 1).
 FT CONFLICT 234 234 L -> I (IN REF. 1).
 FT SEQUENCE 495 AA; 56639 MW; 26AFF6797DDA54D6 CRC64;
 SQ
 Query Match 4.9%; Score 141; DB 1; Length 495;
 Best Local Similarity 21.3%; Pred. No. 0.26;
 Matches 89; Conservative 56; Mismatches 117; Indels 156; Gaps 21;
 QY 52 VPADPNIDAD-LMNIAQDYFNAAVGMDSAVSNKHINILEDNHNADPEYFNKIGPOLTM 110
 DB 163 IDHIEPNDEGIFKIVNDYTG-----EGMN--DQYDELGNFIDYLM 201
 QY 111 DDTIKNSLNHGLSDATNRWG-----LDAIVH-QSLADRENNSTENVIPNYS 156
 DB 202 GENI-DFRNHANTEIKYARVWMEQTQCGFLDAVKHHPAFYKEWIERHVSQVAPRL 260
 QY 157 FVRA-HDNNSQDQIGNAIRDVGKQYHTTFDEQKGDYATQ-----DQNSTVKKYNLYN 211
 DB 261 FIVAEYWSHEVDKLTQYLDQVEGK--TMLF-----DAPLQMKPEASRMGRDYDMTQ 310
 QY 212 I-----PASVAILLTNKDTIPRYVYGLD 234
 DB 311 IFGTGLVEADPFHAYTLVANHDTPQLALEAPVPEWFKPLAYALILLRENGVPSVFPDL 370
 QY 235 YTDGGGYM-----EQQTRYD-----TLTNLLSRVYVAGGQSMOTSVGGNNILTSYR 285
 DB 371 Y--GAHYEDVGGDQYFPIDMPIIEQLDELILARQFANG---VQTLFFDHPNCATFS-- 423
 QY 286 YKGAMTATDTGDETGTQIGVWVNTNPKLGVNDKVVYLVNHAHAAKQYRAAVLTFT 345
 DB 424 -----RSGTDE--FPGCVVVMSEN-----GDDGKTIHLGENYGNKTRDFLGNQ 466
 QY 346 DGVINTSDQAPVAMTDENGDLYLSSHNLVWVKEADTAOGYANPDVSGYLAVW 403
 DB 467 ERVV-----TDENGA-----TFFCNG-----GSVSVW 490

RESULT 13

CDGU_BACCI STANDARD; PRT: 713 AA.

AC P43379;

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 15-JUL-1995 (Rel. 38, Last annotation update)

DE Cyclomaltodextrin glucanotransferase precursor (EC 2.4.1.19)

DE (Cyclodextrin-glycosyltransferase) (CGTase).

OS Bacillus circulans.

OC Bacillus; Firmicutes; Bacillus/Clostridium group;

OC Bacillus/Staphylococcus group; Bacillus.

OX NCBI_Taxid=1397;

RN [1]

RP SEQUENCE FROM N.A., AND X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).

RC STRAIN=251;

RC MEDLINE=94149761; PubMed=8107143;

RA Lawson C.L., van Montfort R., Strokopytov B., Rozeboom H.J.,

RA Kalk K.H., de Vries G.E., Penninga D., Dijkhuizen L., Dijkstra B.W.;

RT "Nucleotide sequence and X-ray structure of cyclodextrin

RT glycosyltransferase from *Bacillus circulans* strain 251 in a maltose-

RT dependent crystal form.";

RL J. Mol. Biol. 236:590-600(1994).

RN [2]

RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS).

RC STRAIN=251;

RC MEDLINE=96094317; PubMed=7493956;

RA Knegtel R.M.A., Strokopytov B., Penninga D., Faber O.G.,

RA Rozeboom H.J., Kalk K.H., Dijkhuizen L., Dijkstra B.W.;

RT "Crystallographic studies of the interaction of cyclodextrin

RT glycosyltransferase from *Bacillus circulans* strain 251 with natural

RT substrates and products.";

RL J. Biol. Chem. 270:29256-29264(1995).

RN [3]

RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).

RC MEDLINE=9711581; PubMed=8955113;

RA Penninga D., van der Veen B.A., Knegtel R.M., van Hijum S.A.,

RA Rozeboom H.J., Kalk K.H., Dijkstra B.W., Dijkhuizen L.;

RT "The raw starch binding domain of cyclodextrin glycosyltransferase

RT from *Bacillus circulans* strain 251.";

RL J. Biol. Chem. 271:32777-32784(1996).

RN [4]

RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS) OF COMPLEX WITH INHIBITOR.

RC STRAIN=251;

RC MEDLINE=96264806; PubMed=8672460;

RA Strokopytov B., Knegtel R.M.A., Penninga D., Rozeboom H.J., Kalk K.H.,

RA Dijkhuizen L., Dijkstra B.W.;

RT "Structure of cyclodextrin glycosyltransferase complexed with a

RT maltotriose inhibitor at 2.6-A resolution. Implications for product

RT specificity.";

RL Biochemistry 35:4241-4249(1996).

CC -|- CATALYTIC ACTIVITY: Degrades starch to cyclodextrins by formation

CC of a 1,4-alpha-D-glucosidic bond.

CC -|- COFACTOR: BINDS TWO CALCIUM IONS.

CC -|- SURUNIT: MONOMER.

CC -|- DOMAIN: MAY CONSIST OF TWO PROTEIN DOMAINS: THE ONE IN THE AMINO-

CC TERMINAL SIDE CLEAVES THE ALPHA-1,4-GLUCOSIDIC BOND IN STARCH, AND

CC THE OTHER IN THE C-TERMINAL SIDE CATALYZES OTHER ACTIVITIES,

CC INCLUDING THE RECONSTITUTION OF AN ALPHA-1,4-GLUCOSIDIC LINKAGE

CC FOR CYCLIZING THE MALTOOLIGOSACCHARIDE PRODUCED.

CC -|- SIMILARITY: BELONGS TO FAMILY 13 OF GLYCOSYL HYDROLASES, ALSO

CC KNOWN AS THE ALPHA-AMYLASE FAMILY.

CC

CC THIS SWISS-PROT entry is copyrighted. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>).

CC or send an email to license@isb-sib.ch).

CC

CC EMBL; X78145; CAA55023.1; -.

CC PDB; 2CXG; 14-OCT-98.

DR PDB; 1CDG; 08-MAR-95.

DR PDB; 1CGV; 27-FEB-95.

DR PDB; 1CGW; 27-FEB-95.

DR PDB; 1CGX; 07-FEB-95.

DR PDB; 1CGY; 07-FEB-95.

DR PDB; 1CXE; 15-DEC-95.

DR PDB; 1CXF; 15-DEC-95.

DR PDB; 1CXH; 15-DEC-95.

DR PDB; 1CXI; 15-DEC-95.

DR PDB; 2DIJ; 13-JAN-99.

DR PDB; 1TCM; 21-APR-97.

DR InterPro; IPR000461; Alpha_amyase.

DR InterPro; IPR002044; CBD_4.

DR InterPro; IPR002909; IPT_TIG.

DR Pfam; PF00128; alpha-amyase; 1.

DR Pfam; PF02806; alpha-amyase_C; 1.

DR Pfam; PF00686; CBD_4; 1.

DR Pfam; PF01833; TIG; 1.

DR ProDom; PD001568; CBD_4; 1.

KW transferase; Glycosyltransferase; Calcium; Signal; 3D-structure.

FT SIGNAL 1 25

FT CHAIN 26 713

FT DOMAIN 28 165

FT DOMAIN 166 229

FT DOMAIN 230 433

FT DOMAIN 434 522

FT DOMAIN 523 609

FT DOMAIN 610 713

FT DISULFID 70 77

FT ACT_SITE 256 256

FT ACT_SITE 284 284

FT ACT_SITE 355 355

SQ SEQUENCE 713 AA; 77309 MW; 8ABBF2C633A004B CRC64;

Query Match 4.9%; Score 139.5; DB 1; Length 713;

Best Local Similarity 20.0%; Pred. No. 0.5;

Matches 110; Conservative 76; Mismatches 212; Indels 151; Gaps 24;

QY 11 ANDIDNSP-----VVOAEQLNWLIIYLLNFG---TITANNDQANF-DSVRVDAPD 56

DB 129 ARDFKTNPAYGTIADFONLIAAHAKNIKVIIDFAPNHTSPASSDQPSFAEMGRLYDNG 188

QY 57 NIDADLMTAODYFNAYGMSDSDAVSN---KHINILEDNNHADPEYFNKIGNPQLTMDDT 113

DB 189 TLGGTYNTDQNLFFHNGTGTFSTTEGNYKLYDLADLNHNS-----TVQVY 237

QY 114 IKNSLNHGLSDATNRWGLDAIVH-----QSLADRENN-----STENV 150

DB 238 LKDAIKRWLDLIGDGRMDAVKHMFGKQKSFMAAVNKKPVTFCEWFLGVNEVSPENH 297

QY 151 VIPNYSFVRAHDNNSODQIQONAIRDVTGKDYHTFTFDEQKIDAYIQDNSTVVKYNYI 210

DB 298 KFEANESGMSILDPRFAQKVRQVFRDNTNNYGLKMLEGSAADYAGVDQVTFIDNHME 357

QY 211 NIPAS-----VAILLTNKDTIPRYYG-DLYTDGGOYMEHTRY-----YD 250

DB 358 RFHANANRKEQLAFLFTLSRG-VPAIIYGEYQMSGGTDPDNNRIPSPFSISTATQ 416

QY 251 TITNLKSR-----VKYVAG-----GQSNQTMMSGVGNNTILTSVRYKGG 289

DB 417 VIQKLAPLRKCPAIAYGSTQERWINNDVLIYERKFGSNVAVAVNRNLNAPASI---SG 473

QY 290 AMTATDTGT-----DETFQIGVVVSNWPNLKLGVNDKVVLEHMAAKHQOQRA 339

DB 474 LVTSLPQGSYNDVLGGLLNGNLTLSVSGGGAAS---NFTLAAGGTAV-----WQVTA 521

QY 340 AVLTITDGVNTSDGQAPVAMDENGDLVLSHNLVNVNGKEADPAVGYANPDVSGY- 398

DB 522 ATATPTIGHVGPMMKPGVTITIDGRG--FGSKGRVYFG---TTAVSG---ADITISWE 572

QY 399 ---LAYVVPVVGASDNDQARTAPSTKNSGNSAYRTNAAFDNSVIFFAFNFVYTPKKESE 455

344 TTDGV---INVTSDGAPVAMTDENGDL-----YLSHNLVNVNGKEADTAVQGYA 391
Db 1396 ILDAVAPTLSDSDTADPVYTDNPFQITGTATDRAQYLS---LAINGSHVASQADIN1 1452
QY 392 NPDVSYLAWVVPVGSNDQARTAPSTKNSCNSAYRTNRAFDNSVIFEAESNFVYPT 451
Db 1453 NSGKPGHMAIDQVKLLGKNVLTVAVTDSENNTTKKITVYIEPKTKTAAAPT---VTPS 1509
QY 452 KESERANVRAQNADEFASIGTSEFEMAPQYNSKDRFTID-----STIDNG---YAPT 502
Db 1510 TTPAKTVILTANA---ATGET-----VQYSADGKTYQDYPAGVTVTANGTEKFKST 1561
QY 503 DRYDLMGSEPN-----KYGTDEDLRINAQAL 528
Db 1562 DLY--GNESPAVDVYVVTNIKADDPQAQLOTAQAL 1593
RESULT 16
VG37_BPT4
ID VG37_BPT4 STANDARD; PRT; 1026 AA.
AC P03744;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Long tail fiber protein p37 (Protein Gp37) (Receptor recognizing protein).
GN 37.
OS Bacteriophage T4.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;
OC T4-like phages
OX NCBI_TaxID=10665;
RN [1]
RP MEDLINE=82170495; PubMed=7338921;
RA Oliver D.B., Crowther R.A.;
RT "DNA sequence of the tail fibre genes 36 and 37 of bacteriophage T4.";
RL J. Mol. Biol. 153:545-568(1991).
RN [2]
RP SEQUENCE FROM N.A.
RA Kutter E., Arisaka F., Kunisawa T., Tsugita A., Mosig G.,
RA Mesyanzhinov V., Ruger W., Stidham T., Thomas E.;
RT Bacteriophage T4 genome analysis;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: STRUCTURAL COMPONENT OF THE DISTAL-HALF TAIL FIBER. IT
CC CONSTITUTES THE PART OF THE LONG TAIL FIBERS THAT RECOGNIZES THE
CC BACTERIAL RECEPTOR. THE TAIL FIBER OF T4 IS ABOUT 1600 ANGSTROMS
CC LONG WITH A KINK IN THE MIDDLE THAT DIVIDES THE FIBER INTO
CC PROXIMAL AND DISTAL HALVES. THE THIN TIP OF THE DISTAL HALF-FIBER
CC INTERACTS WITH THE BACTERIAL LIPOPOLYSACCHARIDE RECEPTOR AND
CC SPECIFIES THE HOST RANGE OF THE PHAGE.
CC -1- SUBUNIT: THE DISTAL HALF-FIBER CONTAINS TWO MOLECULES EACH OF GP36
CC AND GP37 AND ONE MOLECULE OF GP35.
CC -1- MISCELLANEOUS: THE TWO GP37 PROTEIN CHAINS RUN IN PARALLEL, THE
CC LENGTH OF THE DISTAL HALF-FIBER, WITH THE AMINO END NEAR THE
CC CENTER KINK OF THE FIBER AND THE CARBOXYL END AT THE DISTAL TIP.
CC THE OTHER POLYPEPTIDES ARE DISTRIBUTED UNIFORMLY ALONG THE LENGTH
CC OF THE DISTAL HALF-FIBER.
CC -1- SIMILARITY: BELONGS TO THE TAIL FIBER FAMILY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; V00863; CAA24228.1; --
DR EMBL; J02509; AAA32514.1; --
DR EMBL; AF158101; AA042460.1; --
DR PIR; A04369; TLBP74.
KW 'fiber protein.

SQ SEQUENCE 1026 AA; 109224 MW; 965C1F01E308E1BC CRC64;

Query Match 4.8%; Score 136; DB 1; Length 1026;
Best Local Similarity 19.9%; Pred. No. 1.3;
Matches 119; Conservative 66; Mismatches 236; Indels 176; Gaps 27;

QY 36 GTITANNQANFDSVRVDP-----DNIDADLMNIAQDYFNAAAYGMDSDAVSNKHINILED 91
Db 66 GLRLNGDYVQVGMTVNGPIGSTDGVTKIPFRSQGSF---YARATNDTNAHL----- 117
QY 92 W-NHADPEYENKI--GNPOLTMDDTIK-----NSLNHG-----LSD 124
Db 118 WEFENADCTGEGVIYARPQTTTDEIRLVRQGTGTANSEFFRSINGEFGANRILASD 177
QY 125 A--TNRMGLDAIVHOSLADRENNTENVPYIPSVFRAHDNNSQIQAINRDVTKGKYH 182
Db 178 SLVTKRIAVDTVIHDAKAFQGYDHSLV---NYVYPGTGETGVNLYRKVRKAKSGGTIYH 234
QY 183 TTFEDEQKIDAYIODQNSTVKYNLYNIPASAYAILLTNK-----DTIPRVYVGDLY 235
Db 235 EIV--TAQTGLADEVSWMSGDTPVKLYGIRDGDMIRKNSLALGTTFTNFPSSDYGNVG 292
QY 236 TDGGQVMEHQTRYIDTLTNLLKSRV---KYVAGQSMOTMS----- 273
Db 293 VMGDKYLV---LGDVTGLSYKKTGVDLVGGYGVASITPDSEFRSTRKGFGRSEDOG 348
QY 274 ---VGNNNILTSVR-----YKGMATATD----- 295
Db 349 ATWIMPETNAALLSVQTOADNNAAGDQTHIGYNAGGKMNHFFRGTOHNNINTQOGMEIN 408
QY 296 ---TGTDETRTQIGVWVS-----NTPNLKLGVDKV-----VLH 327
Db 409 PGLIKLVTSNNVQFYADGTIISIQIKLDNEIFPKSNRTAGLKFGAPSDQDGTRTIQW 468
QY 328 MGAHNKQYTRAAVLTTDGVINYTSDOGAPVA--MTDENGDIYLSHNLVNVNGKE--- 381
Db 469 NGGTRESQNKVYIIRKAWGNSFNATGDRSREIVFQVSDSQG-YIFYAHRKAPTGDSTIGR 527
QY 382 -EADTAVQGYA-----NPDVSGYLAVNVFVGCASD---QDARTAPSTEKNSSG--NSAY 428
Db 528 IEAQFAGDVYAKGIIANGNFRVYVGSALAGNVTMSNGLFVQGGSSITGVQKIGGTANALR 587
QY 429 RTAAADSNSVIFEAESNFVTPPKESERAN-----VRIAONADFFASLGFSTF 476
Db 588 INNAEYGA-IFRSESNFYIIPINQNEGSDIHSLRPVRIGLN-DGAVGLGRDSF 642

RESULT 17

CDGT_BACSS
ID CDGT_BACSS STANDARD; PRT; 718 AA.
AC P31747;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-JUL-1993 (Rel. 26, Last annotation update)
DE Cyclomaltoextrin glucanotransferase precursor (EC 2.4.1.19)
DE (Cyclodextrin-glycosyltransferase) (CGTase).
GN CGT.
OS Bacillus sp. (strain 6.6.3).
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=29335;
RN [1]
RP SEQUENCE FROM N.A.
RA Akhmetzhanov A.A.;
RL Submitted (MAY-1992) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: Degrades starch to cyclodextrins by formation
CC of a 1,4-alpha-D-glucosidic bond.
CC -1- COFACTOR: BINDS TWO CALCIUM IONS.
CC -1- SUBUNIT: MONOMER.
CC -1- MISCELLANEOUS: CGTASE MAY CONSIST OF TWO PROTEIN DOMAINS: THE ONE
CC IN THE AMINO-TERMINAL SIDE CLEAVES THE ALPHA-1,4-GLUCOSIDIC BOND
CC IN STARCH, AND THE OTHER IN THE C-TERMINAL SIDE CATALYZES OTHER

CC ACTIVITIES, INCLUDING THE RECONSTITUTION OF AN
CC ALPHA-1,4-GLUCOSIDIC LINKAGE FOR CYCLIZING THE
CC MALTOOLIGOSACCHARIDE PRODUCED.
CC -1- SIMILARITY: BELONGS TO FAMILY 13 OF GLYCOSYL HYDROLASES, ALSO
CC KNOWN AS THE ALPHA-AMYLASE FAMILY.

CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: X66106; CAA46901.1; -
CC PIR: S21532; ALBSG6.
CC HSSP: P30920; ICG1.
CC InterPro: IPR000461; Alpha_amylase.
CC InterPro: IPR002044; CBD_4.
CC InterPro: IPR002909; IPT_TIG.
CC Pfam: PF00128; alpha-amylase; 1.
CC Pfam: PF02806; alpha-amylase_C; 1.
CC Pfam: PF00886; CBD_4; 1.
CC Pfam: PF01833; TIG; 1.
CC ProDom: PD001568; CBD_4; 1.
CC TransFam: Glycosyltransferase; Calcium; Signal.
CC SIGNAL 1 34 POTENTIAL.
CC CHAIN 35 718
CC DOMAIN 35 172 A1.
CC DOMAIN 173 236 B.
CC DOMAIN 237 440 A2.
CC DOMAIN 441 526 C.
CC DOMAIN 529 614 D.
CC DOMAIN 615 718 E.
CC DISULFID 77 84 BY SIMILARITY.
CC ACT_SITE 263 263 BY SIMILARITY.
CC ACT_SITE 291 291 BY SIMILARITY.
CC ACT_SITE 362 362 BY SIMILARITY.
CC SEQUENCE 718 AA; 78014 MW; 7644096D402707E5 CRC64;

Query Match 4.7%; Score 134; DB 1; Length 718;
Best Local Similarity 21.6%; Pred. No. 1.1; 177; Indels 140; Gaps 24;
Matches 105; Conservative 65; Mismatches 140; Gaps 24;

QY 64 NIAQYFNAAYGMSDASVN---KHINLEEDNHADPEYFNKIGNPOLTMDYTKNSLNH 120
DB 203 NDTNGYFHNGGSDFSLENGIYKNLYDLADFNNH-----ATIDKYFKDAIKL 251
QY 121 GLSDATNRWGLDAIVHQSADREN-----NSTENVVPIPNYSF 157
DB 252 WLDMGVDSIRVDAYKTHIALGWQKSWMSIYVHKPVFTFGFWFLGSAASADADNTDFANKSG 311
QY 158 VRAHNNNSQIQONAIRDVTGDKYHTFTFEQKIGDAYIQDNQSTVKKNLYN----- 211
DB 312 KSLLDRENSAVRVFNDTNSWY-----ALSDMI---NSTADYNDQNDYTFI 358
QY 212 -----IPASAYAILLNKDIPIRYZG-DLYTDGGQVMEHQ----- 245
DB 359 DNHMDREKTSVANNRRLEQALFTLSRG-VPAIYITGEQYLTGNGDPPDNRAKMPFSK 417
QY 246 -TRYDITNLN--LKSRYKYAGGQSMQMSVGGNNILTSVR-YGK----- 288
DB 418 STTAFNVISKLAPLRKSNPAIAYGTSQORWI---NNDVYTYERFGKSVAVAVNRNLST 474
QY 289 -----GAMTATDGTGDETRTQIGVVSNTNPKLGVNDKVVLMHGAHKKQYRAAVLT 343
DB 475 PANITGLSFLPTGS-YTDVLG-GVLNNGNITSSNGSVNSFTLAAGAT-AVWQYTAETT 531
QY 344 TTDGVINTSDQAPVAMTDENGDLILSSHNLVYNGKEADTAVOGYANPDVSGYLAWV 403
DB 532 PTIGHGVPMKPGNVVITDGRG--FGSTKGTIVFG---TTAVTGAA-----ITSW- 577

QY 404 PVGASDNQDARTAPSTERNKSGNSAYRTNAAFDNSWIFAFSNF-VYTPKESER-----A 457
DB 578 -----EDTQIKVTIPSWA--AGNYAVKVA--NGVNSNAYNHFTILTGDTVTRPVINNA 628
QY 458 NVRIAGN 464
DB 629 STTLGQN 635
RESULT 18
PMFC_PROMI
ID PMFC_PROMI STANDARD; PRT; 828 AA.
AC P53514;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Outer membrane usher protein pmfC precursor.
GN PMFC.
OS Proteus mirabilis.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Proteus.
OX NCBI_TaxID=584;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H14320;
RX MEDLINE=95047519; PubMed=7959033;
RA Massad G., Mobley H.L.T.;
RT "Genetic organization and complete sequence of the Proteus mirabilis
RT pmfC fimbrial operon";
RL Gene 150:101-104(1994).
CC -1- FUNCTION: INVOLVED IN THE EXPORT AND ASSEMBLY OF PMF FIMBRIAL
CC SUBUNIT ACROSS THE OUTER MEMBRANE.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane
CC (By similarity).
CC -1- SIMILARITY: BELONGS TO THE FIMBRIAL EXPORT USHER FAMILY.
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: 235428; CAA84590.1; -
CC InterPro: IPR000015; Fimb_usher.
CC Pfam: PF00877; Usher; 1.
CC PROSITE: PS01151; FIMBRIAL_USHER; 1.
KW Outer membrane; Transmembrane; Fimbria; Transport; Signal.
FT SIGNAL 1 28 POTENTIAL.
FT CHAIN 29 828 OUTER MEMBRANE USHER PROTEIN PMFC.
SQ SEQUENCE 828 AA; 93108 MW; B1B05992A64BA44 CRC64;

Query Match 4.6%; Score 132.5; DB 1; Length 828;
Best Local Similarity 20.8%; Pred. No. 1.6;
Matches 96; Conservative 70; Mismatches 146; Indels 149; Gaps 28;

QY 10 LANDIDNS-PPVQAEQLNWLILNFG-TITANNQANFDSVRVDAPDNIDALMNIQA 67
DB 416 LSGDITQSYKTYDNEKINGMSFKLAKTFDEYHSTTFAGYRFS-----EKTFRFSQ 470
QY 68 DYFNAAYGMSDASVSNKHINLEEDNHADPEYFNKIGNPOLTMDYTKNSLNHGLSDATN 127
DB 471 -YIDERY-----NGIN-----NGYKEMTTITGNKTFWADDAEKSTLY-LS----- 511
QY 128 RWGLDAIVHQSADRENNSSTENVVIP-NYSFVRAHNNNSQIQONAIR-DVTGDKYHTFT 185
DB 512 -----YHQNYWDKNTQEQGYGVTSVNSFSLMGIEQNTN---LSAFTQYKGNFTDLS 562
QY 186 FE-----DEQKIDAYIQDNQSTVKKNLYNIPASAYAILLNKDTIPRVYGDLY----- 235
DB 563 FNISLPLSGSGNIGNLYQDNNGKVTQM-----ASYA---DNRD-----YNNLWIRTAG 607

```
QY 236 -----TDGQYMEHQTTRYDITNLLKSRKVIYVAGQSQMOTMSVGGNNILTSRYG 287
DB 608 LSSDKKANTDG--YQHRSQAEINANASYQODNYLAVGATIK-----GG----FTATRYG 657
QY 288 -----KGMATAT-----DTETRTQIGGVVVS-----NTPNLKLGVDK 323
DB 658 ALHSSSTSSPARIMVDYDGVAGVPFNGQSITTRNFGIGVLTDLTSNNVDARIDV-DK 716
QY 324 VVLHMAAHKNOQYRAAVLT--TDGVINY-----SHNLVNGKEADTAV 387
DB 717 M-----DQDIETRAKIASTTITTEGAIGYQFPVQROGERLMAVLQTTDNKYPFGAEVT 769
QY 353 SDOGAPVAMTDNGDLYLS-----SHNLVNGKEADTAV 387
DB 770 NOKGESIGNVMEGLVYIAGVNLNESLNVWNGKTQCSITI 810

RESULT 19
COLA_CLOPE
ID COLA_CLOPE STANDARD; PRT; 1104 AA.
AC P43153;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NAR-2002 (Rel. 41, Last sequence update)
DT 01-NAR-2002 (Rel. 41, Last annotation update)
DE Microbial collagenase precursor (EC 3.4.24.3) (120 kDa collagenase).
GN COLA OR CP60173.
OS Clostridium perfringens.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1502;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 87-113.
RC STRAIN=NCIB 10662;
RX MEDLINE=94110220; PubMed=8282691;
RA Matsushita O., Yoshihara K., Katayama S.-I., Minami J., Okabe A.;
RT "Purification and characterization of Clostridium perfringens 120-
RT kilodalton collagenase and nucleotide sequence of the corresponding
RT gene.";
RN [2]
RP J. Bacteriol. 176:149-156(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=13 / Type A;
RX PubMed=11792842;
RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
RT "Complete genome sequence of Clostridium perfringens, an anaerobic
RT flesh-eater.";
RN [3]
RP Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
RN [3]
RP SEQUENCE OF 1073-1104 FROM N.A.
RC STRAIN=NCIB 10662;
RA Matsushita O.;
RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: Digestion of native collagen in the triple
CC helical region at Xaa-I-Gly bonds. With synthetic peptides, a
CC preference is shown for Gly at P3 and P1'; Pro and Ala at P2 and
CC P2'; and hydroxyproline, Ala or Arg at P3'.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M9.
CC -1- SIMILARITY: CONTAINS 1 PKD DOMAIN.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D13791; BAA02941.1; ..
CC EMBL; AP003185; BAB79879.1; ..
CC EMBL; D50309; BAA08848.1; ..
CC 1.EROPS; M09.002; ..
```

```
DR InterPro; IPR002169; Micollptase.
DR InterPro; IPR000601; PKD_domain.
DR InterPro; IPR00130; zn_mtpeptidse.
DR Pfam; PF01752; Peptidase_M9; 1.
DR Pfam; PF00801; PKD; 1.
DR PRINTS; PR00931; MICOLLPTASE.
DR SMART; SM00089; PKD; 1.
DR PROSITE; PS50093; PKD; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Hydrolyase; Metalloprotease; zinc; zymogen; signal; complete proteome.
FT SIGNAL 1 39
FT PROPEP 40 86
FT CHAIN 87 1104
FT DOMAIN 774 862
FT METAL 502 503
FT ACT_SITE 506 506
FT METAL 506 506
FT CONFLICT 38 38
FT CONFLICT 722 722
FT CONFLICT 748 748
FT CONFLICT 945 945
FT CONFLICT 970 970
FT CONFLICT 987 987
FT CONFLICT 1098 1098
FT CONFLICT 1104 1104
SQ SEQUENCE 1104 AA; 125935 MW; F4B7377194ED021C CRC64;

Query Match 4.6%; Score 131.5; DB 1; Length 1104;
Best Local Similarity 19.3%; Pred. No. 2.6;
Matches 101; Conservative 81; Mismatches 173; Indels 169; Gaps 26;

QY 68 DYENAAVGMDSDAVSNKHINILEDVFNHADPEVFNIGNPQLTMDTITKNSLNHGLSD--- 124
DB 581 DFINGVGFAL-SNYMTNNMGM-----FNK-----MTYIKNNDSVGYKDYIA 621
QY 125 -ATNRWGLDAIYHQSLADRENSTENVIPNTS--FYRAHDNNSQDIQNAIRDVYTGKYD 181
DB 622 SHSSDYGLND-NYQYVMDSLNINIDNLDVPLVSDVYNGCHEAKDINEITNDIKEYSN--- 677
QY 182 HTFTFEDEOKGIDAVIQOONSTVKK---YNLNIIPASTAAILLTNKDTPRVVYGLDYDG 238
DB 678 -----IKDLSSNVKESQFFTYDMRGTV-----G 702
QY 239 GOYMEHQTTRYD---TLTNLLKSRKVIYVAGQSQMOTM-----SVGNNK-ILTSVRYGKG 289
DB 703 GRSGQGENDKDNKLNLDLLELSKSNWGYTYTAYFVNKHVYDNGNGYVYDVYVFGHGN 762
QY 290 AMTATDTGTDETRTOGIGVYVNTPNLKLGVNDKVVHL----- 327
DB 763 ----TDTNFD-----VHVNKEPKAVIKSDSSVIEEINFDGTEKDEDEGETKXAEW 810
QY 328 -MGAHKNQYRAAVLTITTDGVINYNTSDQAPVAMTDENGDLVLSHNL-VYNGKEADT 385
DB 811 DFDGGEKSNKAKAT-----HKYNTGEVLTVDNNGGINTESKKIKVYEDK----- 859
QY 386 AVQGYANPOVSGYLVAVVVPVAGSDNQDAPTSKNSGNSAYRTNAAFDSNVIFEAFSN 445
DB 860 -----PVEVINESE-----PNNDPEKANQIAKSNLWVKGTLSEEDYS 897
QY 446 FVYTPPKESRANVRITQADNADFASLGT-----SFEAFQYNSKQRTFL--DSTID 496
DB 898 KYVFDV--AKKGNVKITLNN--LNSVGITWLYKEGDLNNVLYATGNDGTVLKGEKTL 953
QY 497 NGVAFYDVRDIGH-SEPNKYGT-----DEDLRNAIQALHKAGIQ 534
DB 954 PG-----RYLVSVITYDNGSGTYVNVNKNLXNEVKETAIDAIK 992

RESULT 20
APU_THESA
ID APU_THESA STANDARD; PRT; 1279 AA.
AC P36905;
DT 01-JUN-1994 (Rel. 29, Created)
```

DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DE 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Amylopullulanase precursor (Alpha-amylose/pullulanase) [Includes:
 DE Alpha-amylose (EC 3.2.1.1) (1,4-alpha-D-glucan glucanohydrolase);
 DE Pullulanase (EC 3.2.1.41) (1,4-alpha-D-glucan glucanohydrolase)
 DE (Alpha-dextrin endo-1,6-alpha-glucosidase)]
 GN APF.
 OS Thermoanaerobacter saccharolyticum.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Thermoanaerobacter group; Thermoanaerobacterium.
 OX NCBI_TaxID=28896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=B6A-RI;
 RX MEDLINE=9416125; PubMed=8117096;
 RA Ramesh M.V., Podkovyrov S.M., Lowe S.E., Zeikus J.G.;
 RT "Cloning and sequencing of the Thermoanaerobacterium saccharolyticum
 RT B6A-RI apu gene and purification and characterization of the
 RT amylopullulanase from Escherichia coli.";
 RL Appl. Environ. Microbiol. 60:94-101(1994).
 RN [2]
 RP IDENTIFICATION OF PROBABLE VECTOR CONTAMINATION.
 RA Robinson K.;
 CC Unpublished observations (NOV-1994).
 CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-alpha-glucosidic
 CC linkages in oligosaccharides and polysaccharides.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of (1->6)-alpha-D-glucosidic
 CC linkages in pullulan and in amylopectin and glycogen, and the
 CC alpha- and beta-limit dextrins of amylopectin and glycogen.
 CC -1- SIMILARITY: BELONGS TO FAMILY 13 OF GLYCOSYL HYDROLASES, ALSO
 CC KNOWN AS THE ALPHA-AMYLASE FAMILY.
 CC -1- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
 CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN BY THE PRESENCE OF
 CC AN EXTRA C-TERMINAL SEGMENT OF 9 RESIDUES THAT SEEMS TO ORIGINATE
 CC FROM A PUC-TYPE VECTOR.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: L07762; AA019800.1; ALT_SEQ.
 DR HSSP: Q08751; 1BVZ.
 DR InterPro: IPR000461; Alpha-amylose.
 DR InterPro: IPR003961; FN.III.
 DR InterPro: IPR004185; alpha-amylose_N.
 DR Pfam: PF00128; alpha-amylose.1.
 DR Pfam: PF02806; alpha-amylose.C; 1.
 DR Pfam: PF02903; alpha-amylose_N; 1.
 DR Pfam: PF00041; fn3; 2.
 DR SMART: SM00060; FN3; 2.
 KW Hydrolase; Glycosidase; Carbohydrate metabolism; Signal; Repeat;
 KW Multifunctional enzyme.
 FT SIGNAL 1 35 POTENTIAL.
 FT CHAIN 36 1279 AMYLOPULLULANASE.
 FT DOMAIN 929 1017 FIBRONECTIN TYPE-III 1.
 FT DOMAIN 1156 1248 FIBRONECTIN TYPE-III 2.
 FT ACT_SITE 629 629 BY SIMILARITY.
 FT ACT_SITE 658 658 BY SIMILARITY.
 FT ACT_SITE 735 735 BY SIMILARITY.
 SQ SEQUENCE 1279 AA; 142430 MW; 095CCBCA391624DD CRC64;

 Query Match 4.58; Score 131; DB 1; Length 1279;
 Best Local Similarity 20.28; Pred. No. 3.4;
 Matches 122; Conservative 83; Mismatches 202; Indels 196; Gaps 35;

 QY 40 ANNOANFDSVRYDAPDNIDALMNIADYFNAYGMSDAVSNKHNILBD--WNHA-- 95
 DB 483 SNHKYDTADYTKID-----EMGTTDDFEK----LMSDAHA-KGINILDGVENHTSD 530

QY 96 DPEYENKIGN-POLTMDDTYKNSLNHGL-----SDATNR--WGDLAI-VQSLAD 141
 DB 531 DSIYENRYGKYPDLGAYQDWKDG--NOSLSPYGDMWTINSDETGYECWGYDSLPIVKSNG 589
 QY 142 RENNSTE--NVVPIYNSFVRAH-----DNNSOD-----QIONAIR----- 174
 DB 590 SEYNTSWANFIINDKNKAIKWLNPDENLNDGADWRDLVENEVAHFWTHFRDAINTV 649
 QY 175 -----DVTGKDYHTF-----TFED-----EQKGIDAYIQD 199
 DB 650 KPEAFPIAENMGDASLDLLGDSFNSVMYQFRNDIIDFLIGQSFDDGNGHNPIDAAKLD 709
 QY 200 QNSTYKKNLYNIPASYAI--LLTNKDTIPRYYGDLTYDGGYMEHQTRYDYDTLNLK 257
 DB 710 QR-LMSIYERYPLPAPYSTPNLLGSHDTM-----SILTVFGYNSADPNENSDAARLAE 762
 QY 258 SRVYVAGSGMOTMSVGGNNILATSVRYGKGAMTATDTGTDETRT----- 303
 DB 763 QKLKLA---TILQMGYFG-----MADIFYGDEAGYSGGKDPDRRTFFWGNEDTALQDFP 814
 QY 304 QGIGVVVSNTPMLKLGVDNKVVLHMGAAHKNQYRAA--VLFTTGVVINYTSDQGAPVAM 361
 DB 815 KNVSSIRNNQVLTG--DLETLA---QNDVYAIGRRIINGKDAFGNSYDPSAAIVAI 868
 QY 362 TDENGDIYLS-----SHNLVYNGKEADTAVQGYAN---PDYSGYLAVWVPVGA 407
 DB 869 NRSNSDQQTITDTKPLRGVAFKDLING--DKSYTINGGQITINIPAMSGVMLI-----S 922
 QY 408 SNQDARTAPSTEKN---SGN-----SAYRMAAFDSNVIFEAFSNFYVYPTKESERAN 458
 DB 923 DQGQDL-TAPQVPSNVVATSGNKVDLSQSDGATGYNIYRSSVEGGLYE----- 972
 QY 459 VRIQNAQDFASLGFTSFEMAPQYNSKDRTELDSTIDMGYAFTRDYDLGMSEPNKYGD 518
 DB 973 -KIASNV-----TGTFEDTNTVINGL-----KYVYIAISAVDELNGE 1007
 QY 519 EDL 521
 DB 1008 SEM 1010
 RESULT 21
 YF06_MYCPN STANDARD; PRT; 793 AA.
 AC F75280;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DE Hypothetical lipoprotein MPN506 precursor (P02_orf793).
 GN MPN506 OR MP336.
 OS Mycoplasma pneumoniae.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
 OC Mycoplasmataceae; Mycoplasma.
 OX NCBI_TaxID=2104;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 29342 / M129;
 RX MEDLINE=97105885; PubMed=8948633;
 RA Himmelfreich R., Hilbert H., Plagens H., Firkel E., Li B.-C.,
 RA Herrmann R.;
 RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
 RT pneumoniae";
 RL Nucleic Acids Res. 24:4420-4449(1996).
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
 CC (potential).
 CC -1- SIMILARITY: BELONGS TO THE MG185 / MG260 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

QY 334 NQYRAAVLTDTGGINVYNTSDQAPVAM 361
||:| : ||| : ||:| :
Db 465 NQKY----VYNTDGTILKVVITGFPVIL 488

RESULT 23

CSG_HALHA STANDARD; PRT; 852 AA.
AC P08198;

DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cell surface glycoprotein precursor (CSG).

OS Halobacterium halobium.
OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;
OC Halobacterium.

OX NCBI_TaxID=2242;
RN [1]

RP SEQUENCE FROM N.A.
RX MEDLINE=87250642; PubMed=3036870;

RA Lechner J., Wieland F.;
RT "The primary structure of a prokaryotic glycoprotein. Cloning and
sequencing of the cell surface glycoprotein gene of halobacteria."

RL J. Biol. Chem. 262:9724-9729(1987).
RN [2]

RP REVIEW.
RX MEDLINE=89372794; PubMed=2673008;

RA Lechner J., Wieland F.;
RT "Structure and biosynthesis of prokaryotic glycoproteins.";

RL Anna. Rev. Biochem. 58:173-194(1989).
CC -!- FUNCTION: THE S-LAYER IS A PARACRYSTALLINE MONO-LAYERED ASSEMBLY

CC OF PROTEINS WHICH COAT THE SURFACE OF THE CELL.
CC -!- SUBCELLULAR LOCATION: CELL WALL. THIS ARCHAEA IS COVERED BY A

CC S-LAYER WITH HEXAGONAL SYMMETRY.
CC -!- PPM: N-LINKED GLYCANS AT ASN-36 CONSISTS OF A GLYCOSAMINOGLYCAN

CC CHAIN CONSTRUCTED BY A REPEATING SULFATED PENTASACCHARIDE BLOCK;
CC OTHER N-LINKED GLYCANS CONTAIN GLC, GLCA AND IDCA; O-LINKED

CC GLYCANS CONSIST OF GLC-GAL DISACCHARIDES.
CC -!- SIMILARITY: TO H.VOLCANII CELL SURFACE GLYCOPROTEIN.

CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>
or send an email to license@isb-sib.ch).

CC
CC EMBL: J02767; AAA72185.1; -.
CC PIR: A28459; A28459.

CC Glycoprotein; Transmembrane; Cell wall; S-layer; Signal.
CC SIGNAL 1 34

CC CHAIN 35 852
CC DOMAIN 789 813
CC TRANSMEM 829 849

CC POTENTIAL.
CC N-LINKED (GLNAC. . .).
CC CARBOHYD 339 339

CC CARBOHYD 398 398
CC CARBOHYD 438 438
CC CARBOHYD 513 513

CC CARBOHYD 513 513
CC CARBOHYD 643 643
CC CARBOHYD 727 727

CC CARBOHYD 751 751
CC CARBOHYD 787 787
CC CARBOHYD 789 789

CC CARBOHYD 791 791
CC CARBOHYD 792 792
CC CARBOHYD 793 793

CC CARBOHYD 793 793
CC CARBOHYD 795 795
CC CARBOHYD 797 797

CC CARBOHYD 798 798
CC CARBOHYD 799 799
CC CARBOHYD 799 799

FT CARBOHYD 801 801 O-LINKED (GAL. . .).
FT CARBOHYD 802 802 O-LINKED (GAL. . .).
FT CARBOHYD 803 803 O-LINKED (GAL. . .).
FT CARBOHYD 806 806 O-LINKED (GAL. . .).
FT CARBOHYD 807 807 O-LINKED (GAL. . .).
FT CARBOHYD 808 808 O-LINKED (GAL. . .).
FT CARBOHYD 811 811 N-LINKED (GLC. . .).
FT CARBOHYD 812 812 O-LINKED (GAL. . .).
FT CARBOHYD 813 813 O-LINKED (GAL. . .).
FT CARBOHYD 815 815 N-LINKED (GLC. . .).
SQ SEQUENCE 852 AA; 89813 MW; 81F0DDF6F1C11505 CRC64;

Query Match 4.5%; Score 129; DB 1; Length 852;

Best Local Similarity 20.3%; Pred. No. 2.6; 240; Indels 178; Gaps 32;
Matches 129; Conservative 87; Mismatches 87; 240; Indels 178; Gaps 32;

QY 3 YRQEFLLANDIIONSPV-----VQAEQLNMLYLLNFGTITAN- 41
||:| : ||| : ||:| :
Db 74 FQGEEDVTFKLDNEKESVATLSRTGSGDEGVLPQMPEDQ-----STGSDSNG 125
QY 42 --NQANFD-----SVRVDPADNTADIMNIAQDYFNAAAGMDSDAVSNKHINI 86
||:| : ||| : ||:| :
Db 126 PDNDEADFGTVQSPSVTLMLEVRNADNDVTGVLATQDDEFSIAVDYNYAAEDLELAV 185
QY 89 LEDWNHAD-----PEYFNKIGN--PQLTMDDTIKNSLNHGLSDATNRWGLDAI 134
||:| : ||| : ||:| :
Db 186 -EDEGLDVTDEILANDQSGAYEDGTGNKGPWTLRFIDIPNNVDAG--DYT-----V 235
QY 135 VHOSLADRE--NNSTENVYIPNYSFVRADHNNSSDQI-----QNAIRDVTGDKYHFTT-- 185
||:| : ||| : ||:| :
Db 236 SVGEVDLDGDATEASAVTISNNKASLNAEDEVVOGANLKYTIENSPGNYHVTID 295
QY 186 ---FEDQKQIDA-----YIQQNSTVKYKMLY-----NIPASYAILLTN 222
||:| : ||| : ||:| :
Db 296 SSOPSSSGADAAKVMRSVGDVYDTGLVVDNDSTTEIVDDYENTSISSDYAYAIWEID 355
QY 223 KDTIPRVYGLYTDGQYMEROTRYDITLNLKSKRVYVAGSQMOTMSVGGNNILIT 282
||:| : ||| : ||:| :
Db 356 -----DGNVGSIEFYLDSS-----ADIDLYPASDTEADPDYVNSNEELT 397
QY 283 SVRYGKAMTATDGTGDETR-----TQIGGVVVSNTPLNKLGVNDKVVLMHGA 330
||:| : ||| : ||:| :
Db 398 NGSALDGVSTDDTDFDVTQCDITLNDPTGAYVVGSEVDINGTAN---EGTDVVLY--- 451
QY 331 AHKQOQYRAAVLTDTG--VINYTSD-----QGAPEVMTDENGEDLYL---SSHLNVNGK 380
||:| : ||| : ||:| :
Db 452 ARDNDFE---LVTVDDGEKSEIYVSDDTFEEDITLSDGDKGDDILGLFCTYRLGLIAK 508
QY 381 BEADTAVQGYA-NPDVSGYLAVVPVGVASDNQDARTAPSTEKNSGNSAYRTNAAFDSDNVI 439
||:| : ||| : ||:| :
Db 509 SDAYNSSGVKDNIDTSDF-----NQGVSTSSSTSTSTVY-DYELTASFETYNGQVADDNQI 562
QY 440 FEAFSNFYVTTKESERANVRIAQNAOFFASLGFTSPFEMAPQYNSSKDRTFDSTID--- 496
||:| : ||| : ||:| :
Db 563 DVEGT---APGKDNVAII-----IGSRGKVKFQ---SISVSDSDTDFEEDIIDISE 607
QY 497 --NGYAFDTRDYLGNSEPNKYGTDLEDLNAIQAL 528
||:| : ||| : ||:| :
Db 608 LRQGSA--SAHILSSGRDGKFG--EDTANSISDL 637

RESULT 24

APU_THETHY

ID APU_THETHY STANDARD; PRT; 1475 AA.

AC P16950;

DT 01-AUG-1990 (Rel. 15, Created)

DT 01-AUG-1990 (Rel. 15, Last sequence update)

DT 01-MAR-2002 (Rel. 41, Last annotation update)

DE Amylopullulanase precursor (Alpha-amylase/pullulanase) [Includes:

DE Alpha-amylase (EC 3.2.1.1) (1,4-alpha-D-glucan glucanohydrolase);

DE Pullulanase (EC 3.2.1.41) (1,4-alpha-D-glucan glucanohydrolase)

DE (Alpha-dextrin endo-1,6-alpha-glucosidase)].

GN apu.
OS thermoaerobacter thermohydrosulfuricus (Clostridium
OC thermohydrosulfuricum).
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OX Thermoanaerobacter group; Thermoanaerobacter.
OX NCBI_TaxID:1516;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-EL01-69;
RX MEDLINE-90362027; PubMed-2391488;
RA Melasniemi H., Palohelmo M., Hemioe L.;
RT "Nucleotide sequence of the alpha-amylase-pullulanase gene from
RT Clostridium thermohydrosulfuricum.";
RL J. Gen. Microbiol. 136:447-454(1990).
RN [2]
RP SEQUENCE OF 32-39.
RC STRAIN-EL01-69;
RX MEDLINE-88268757; PubMed-3260488;
RA Melasniemi H.;
RT "Purification and some properties of the extracellular alpha-amylase-
RT pullulanase produced by Clostridium thermohydrosulfuricum.";
RL Biochem. J. 250:813-818(1988).
RN [3]
RP FIBRONECTIN TYPE III DOMAINS.
RX MEDLINE-93028390; PubMed-1409594;
RA Bork P., Doolittle R.F.;
RT "Proposed acquisition of an animal protein domain by bacteria.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:8990-8994(1992).
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-alpha-glucosidic
CC linkages in oligosaccharides and polysaccharides.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of (1->6)-alpha-D-glucosidic
CC linkages in pullulan and in amylopectin and glycogen, and the
CC alpha- and beta-limit dextrins of amylopectin and glycogen.
CC -1- SIMILARITY: BELONGS TO FAMILY 13 OF GLYCOSYL HYDROLASES, ALSO
CC KNOWN AS THE ALPHA-AMYLASE FAMILY.
CC -1- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M28471; AAA23205.1; -.
DR PIR: A44765; A44765.
DR HSP: Q08751; 18VZ.
DR InterPro: IPR000461; Alpha_amylase.
DR InterPro: IPR003961; FN_III.
DR InterPro: IPR004185; alpha-amylase_N.
DR Pfam: PF00128; alpha-amylase_1.
DR Pfam: PF02806; alpha-amylase_C; 1.
DR Pfam: PF02903; alpha-amylase_N; 1.
DR Pfam: PF00041; fn3; 2.
DR SMART: SM00060; FN3; 2.
KW Hydrolase; Glycosidase; Carbohydrate metabolism; Signal; Repeat;
KW Multifunctional enzyme.
FT SIGNAL 1 31
FT CHAIN 32 1475 AMYLOPULLULANASE.
FT DOMAIN 927 1016 FIBRONECTIN TYPE-III 1.
FT DOMAIN 1163 1255 FIBRONECTIN TYPE-III 2.
FT ACT_SITE 629 629 BY SIMILARITY.
FT ACT_SITE 658 658 BY SIMILARITY.
FT ACT_SITE 735 735 BY SIMILARITY.
SQ SEQUENCE 1475 AA; 165631 MW; 3476C414110B376 CRC64;

Query Match 4.5%; Score 128.5; DB 1; Length 1475;
Best Local Similarity 19.8%; Pred. No. 5.7;
Matches 124; Conservative 79; Mismatches 223; Indels 199; Gaps 30;
QY * 16 NSNPVVAEQEN-----WLYYLLNFGTITANNQANFDSVRVDAPDNIDAD----- 61

Db 579 DSLPVIR--QINGSEYRKSWADFIIN-----NPNASKYWLNPDKNGVADGWRLDV 630
QY 62 LMNIADQFNAAAYGSDAVSNKHLNILLQWNNHADPEYFNKIGNPQLTMDDTIKKSLNKG 121
Db 631 ANEVADFVWFHFGAINTVKPNAPM-VAENWHDASLD-----LLGDSFNSVMNYL 679
QY 122 LSDATNRWGLDA-----IVHQ-----SLADRENNSTENVIPNY-----SFVRAHDN--- 163
Db 680 FRNAVIDFDLKSFDGNGVNVHPIDAAKLDLQRLMSIYERYPLPVEYSTWNLLGSHUTMRI 739
QY 164 -----NSQDQTON-AIRDVT-----GKD-----YH 182
Db 740 LTVFGYNSADENQNSQAAXDLAVKRLKLAAILQMGYPGMPSTIYYGDEAGSGGKDPDNR 799
QY 183 TPTFEDEQKIDAY-----IQDNSTYKKNLYNTPAS---YAI---LLTNKNDTIPRY 230
Db 800 TFWGREDTDLQTFKKNVIRNENOVLTGDELYANGDYAFGRRIINGKDTFGKSY 859
QY 231 YGLDXT-----DGGQYMEHQTRYVD---TLTNLKSRYKYVAGGQSMQMSVGGNNIL 281
Db 860 PDSVAIWINKGDAKQVSITDTTKFIRDGVAFDALSGKTYTVODGK-----IV 907
QY 282 TSVRYCKGAMTADTGTDETRFGIGVYVSTNPNKLGVDNVKVLHMGAAHKNQOYRAAV 341
Db 908 VEVGSDMGAILSDTGNLTAPQPI-----TDLKAVSGNGKVDLSWVYVDKAVSYNIYR 961
QY 342 LTTTDXGVINTSDQGAAPVAMTD---ENGDLVLSHNLVYN-KKEEA----- 383
Db 962 STVKGGLYEKIASNVTQIYTTDETVNGLKYVAVTAVDNDGNSALSNEVAYPAFFPG 1021
QY 384 -----DTAVOGYANPDVSGYLAVVWPGASDNDARTAPSTERNKNS-GNSAYRTNAAF 434
Db 1022 WAGNMNQVTHVIGVNNP-VEVYAEVWAO-GLTDK-----PGQGENMIAOLGYR---Y 1069
QY 435 DSNVIEAFSNFYTTKESERANVRINQNADEFASLGTSFEMAQYNSSDKDTFLDST 494
Db 1070 IGDVGDVNAVYNKVEGVE-----ISKDWTWVDAQ 1101
QY 495 IDNGYAFTRDYLGMSEPNKYGTDE 519
Db 1102 YVGDGNNQKY-NAKFVFDVGVGTWE 1125
RESULT 25
AMY_BACST
ID AMY_BACST STANDARD; PRT; 549 AA.
AC P06279; Q45519;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Alpha-amylase precursor (BC 3.2.1.1) (1,4-alpha-D-glucan
DE glucanohydrolase).
GN AMYS.
OS Bacillus stearothermophilus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Geobacillus.
OX NCBI_TaxID:1422;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 35-39.
RX MEDLINE-85234394; PubMed-3924897;
RA Nakajima R., Imanaka T., Alba S.;
RT "Nucleotide sequence of the Bacillus stearothermophilus alpha-amylase
RT gene.";
RL J. Bacteriol. 163:401-406(1985).
RN [2]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN-D15/PHI300;
RX MEDLINE-86008166; PubMed-3876333;
RA Ihara H., Sasaki T., Tsuboi A., Yamagata H., Tsukagoshi N., Uda S.;
RT "Complete nucleotide sequence of a thermophilic alpha-amylase gene:
RT homology between prokaryotic and eukaryotic alpha-amylases at the
RT active sites.";

J. Biochem. 98:95-103(1985).
[3]
SEQUENCE FROM N.A.
RC STRAIN-NZ-3;
RX MEDLINE-86195857; PubMed-3009417;
RA Gray G.L., Mainzer S.E., Rey M.W., Lamsa M.H., Kindle K.L.,
RA Carmona C., Requaet C.;
RT "Structural genes encoding the thermophilic alpha-amylases of
RT Bacillus stearothermophilus and Bacillus licheniformis.";
RL J. Bacteriol. 166:635-643(1986).
[4]
SEQUENCE FROM N.A.
RP Suominen I., Kaip M., Lautano J., Knowles J., Mantsaelae P.;
RA "Thermostable alpha amylase of Bacillus stearothermophilus: cloning,
RT expression, and secretion by Escherichia coli.";
RL (In) Chaloupka J., Krumphanzl V. (eds.);
RL Extracellular enzymes of microorganisms, pp.129-137, Plenum Press,
RL New York (1987).
[5]
SEQUENCE OF 1-122 FROM N.A., AND SEQUENCE OF 35-48.
RP STRAIN-DY-5;
RC MEDLINE-86059211; PubMed-2999073;
RA Tsukagoshi N., Iritani S., Sasaki T., Takemura T., Ihara H.,
RA Idota Y., Yamagata H., Uda S.;
RT "Efficient synthesis and secretion of a thermophilic alpha-amylase by
RT protein-producing Bacillus brevis 47 carrying the Bacillus
RT stearothermophilus amylase gene.";
RL J. Bacteriol. 164:1182-1187(1985).
CC -I- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-alpha-glucosidic
CC linkages in oligosaccharides and polysaccharides.
CC -I- SIMILARITY: BELONGS TO FAMILY 13 OF GLYCOSYL HYDROLASES, ALSO
CC KNOWN AS THE ALPHA-AMYLASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; M11450; AAA22235.2; -;
DR EMBL; X02769; CAA26547.1; -;
DR EMBL; M57457; AAA22227.1; -;
DR EMBL; M13255; AAA22241.1; -;
DR PIR; A00845; ALBSF.
DR PIR; A24549; A24549.
DR HGSP; P06278; 1VJS.
DR InterPro: IPR000461; Alpha-amylase.
DR Pfam; PF00128; alpha-amylase; 1.
KW Hydrolase; Glycosidase; Carbohydrate metabolism; Signal.
FT SIGNAL 1 34
FT CHAIN 35 549 ALPHA-AMYLASE.
FT ACT_SITE 268 268 BY SIMILARITY.
FT ACT_SITE 272 272 BY SIMILARITY.
FT ACT_SITE 365 365 BY SIMILARITY.
FT CONFLICT 13 13 M -> V (IN REF. 3).
FT CONFLICT 19 19 L -> W (IN REF. 3).
FT CONFLICT 23 23 L -> S (IN REF. 2 AND 3).
FT CONFLICT 31 31 P -> H (IN REF. 2 AND 5).
FT CONFLICT 107 107 A -> T (IN REF. 2 AND 3).
FT CONFLICT 167 167 T -> I (IN REF. 4).
FT CONFLICT 179 179 P -> N (IN REF. 3).
FT CONFLICT 251 251 S -> N (IN REF. 2, 3 AND 4).
FT CONFLICT 260 262 TNI -> RTL (IN REF. 4).
FT CONFLICT 284 284 D -> Y (IN REF. 2, 3 AND 4).
FT CONFLICT 312 312 M -> T (IN REF. 2 AND 3).
FT CONFLICT 338 338 T -> A (IN REF. 2 AND 3).
FT CONFLICT 342 342 R -> S (IN REF. 3).
FT CONFLICT 346 346 T -> N (IN REF. 3).
FT CONFLICT 376 376 V -> C (IN REF. 2 AND 3).
FT CONFLICT 526 527 WS -> RP (IN REF. 2).
FT CONFLICT 527 527 S -> P (IN REF. 2).

FT CONFLICT 535 535 D -> G (IN REF. 2 AND 3).
SQ SEQUENCE 549 AA; 62670 MW; 3A2DD93A955E79D3 CRC64;

Query Match 4.5%; Score 128; DB 1; Length 549;
Best Local Similarity 38.6%; Pred. No. 1.6;
Matches 34; Conservative 9; Mismatches 33; Indels 12; Gaps 2;

QY 460 RTAQNADEFFASLGFTSFEMAPQYNSSKDRTFDLDSTIDNGYARTDRYDLGM-----SEPNK 514
DB 59 KVANEANNLSLGLTALWLPAYKGT-----SRSDVGYGVYDLDLGEFNQKGAVRTK 111

QY 515 YGTDEDLRNAIQALHKAQGVNADWVPD 542
DB 112 YGKQAYLQAIQAHAHAGMQVYADVVD 139

Search completed: August 12, 2002, 08:04:11
Job time: 228 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 12, 2002, 07:59:48 ; Search time 30.83 seconds

(without alignments)
3046.910 Million cell updates/sec

Title: US-09-995-749a-2_copy_972_1514

Perfect score: 2852

Sequence: 1 QTYRGQEFLLANDINSPV.....AIQALHKAGLQVMDWVFDQ 543

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

SPRMBL_19:*

1: sp_archea:*

2: sp_bacteria:*

3: sp_fungi:*

4: sp_human:*

5: sp_invertebrate:*

6: sp_mammal:*

7: sp_mhc:*

8: sp_organelle:*

9: sp_phase:*

10: sp_plant:*

11: sp_rodent:*

12: sp_virus:*

13: sp_vertebrate:*

14: sp_unclassified:*

15: sp_rvirus:*

16: sp_bacteriap:*

17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1539	54.0	1016	2 Q9LCJ7	Q9LCJ7 leuconostoc
2	1532.5	53.7	2057	2 Q9RE05	Q9RE05 leuconostoc
3	1523	53.4	1527	2 Q9ZAR4	Q9ZAR4 leuconostoc
4	1469.5	51.5	1508	2 Q9ZRH5	Q9ZRH5 leuconostoc
5	1467.5	51.5	1477	2 Q9L466	Q9L466 leuconostoc
6	1466.5	51.4	1455	2 Q9S388	Q9S388 streptococc
7	1464.5	51.3	1508	2 Q52224	Q52224 leuconostoc
8	1458.5	51.1	1455	2 Q69397	Q69397 streptococc
9	1453.5	51.0	1455	2 Q69382	Q69382 streptococc
10	1452.5	50.9	1390	2 Q69385	Q69385 streptococc
11	1451.5	50.9	1455	2 Q69391	Q69391 streptococc
12	1414	49.6	1590	2 Q59983	Q59983 streptococc
13	1409	49.4	1590	2 Q52633	Q52633 streptococc
14	1407.5	49.4	1449	2 Q68542	Q68542 streptococc
15	1407.5	49.4	1449	2 Q5264	Q5264 streptococc
16	1383	48.5	1290	2 Q48756	Q48756 leuconostoc

17	1351.5	47.4	1575	2 Q9LCH3	Q9LCH3 streptococc
18	1347.5	47.2	1577	2 Q54178	Q54178 streptococc
19	1288.5	45.2	1577	2 Q55265	Q55265 streptococc
20	1249.5	43.8	1338	2 Q9WXJ4	Q9WXJ4 streptococc
21	1231.5	43.2	1518	2 Q00600	Q00600 streptococc
22	1219	42.7	1512	2 Q9WXJ5	Q9WXJ5 streptococc
23	1179.5	41.4	1599	2 Q00599	Q00599 streptococc
24	169	5.9	501	2 Q93148	Q93148 bacillus sp
25	168	5.9	486	2 Q86875	Q86875 streptococc
26	163.5	5.7	912	2 Q06915	Q06915 lactobacill
27	160.5	5.6	1385	16 Q99W47	Q99W47 ataphylococ
28	157.5	5.5	516	2 Q82839	Q82839 bacillus sp
29	156	5.5	953	2 Q48502	Q48502 lactobacill
30	154.5	5.4	533	2 Q9A054	Q9A054 bacillus me
31	153	5.4	910	9 Q37959	Q37959 lactococcus
32	152.5	5.3	1315	2 Q86488	Q86488 staphylococ
33	151.5	5.3	601	2 Q45821	Q45821 clostridium
34	150	5.3	1377	16 Q98Q28	Q98Q28 mycoplasma
35	149	5.2	1733	2 Q9K114	Q9K114 staphylococ
36	147	5.2	4152	2 Q9ZHL3	Q9ZHL3 haemophilus
37	145.5	5.1	1148	2 Q60045	Q60045 thermococcus
38	144	5.0	429	10 Q40016	Q40016 hordeum vul
39	144	5.0	1305	2 Q9ZE7	Q9ZE7 escherichia
40	144	5.0	6713	16 Q99U54	Q99U54 staphylococ
41	144	5.0	6713	16 Q931R6	Q931R6 staphylococ
42	143.5	5.0	519	2 Q9RQT8	Q9RQT8 cytophaga s
43	143.5	5.0	901	2 Q9XCV8	Q9XCV8 lactobacill
44	143.5	5.0	1306	2 F77070	F77070 escherichia
45	143	5.0	584	16 Q9ZD08	Q9ZD08 rickettsia
46	143	5.0	1181	5 Q97260	Q97260 plasmodium
47	141.5	5.0	786	5 Q77357	Q77357 plasmodium
48	140	4.9	538	2 Q86030	Q86030 vibrio chol
49	139	4.9	484	2 Q50583	Q50583 streptococc
50	139	4.9	634	16 Q97H03	Q97H03 clostridium
51	139	4.9	3890	16 Q99U53	Q99U53 staphylococ
52	136	4.8	1139	5 Q97237	Q97237 plasmodium
53	135.5	4.8	1260	16 Q931M2	Q931M2 staphylococ
54	135	4.7	1578	16 Q92E25	Q92E25 listeria in
55	134.5	4.7	461	2 Q46186	Q46186 clostridium
56	134.5	4.7	1335	2 Q33635	Q33635 staphylococ
57	134	4.7	759	2 Q49099	Q49099 moraxella c
58	134	4.7	2402	2 Q9AER7	Q9AER7 staphylococ
59	133.5	4.7	953	16 Q99W48	Q99W48 staphylococ
60	133.5	4.7	1021	2 Q46085	Q46085 clostridium
61	133.5	4.7	1072	16 Q9CF64	Q9CF64 lactococcus
62	133.5	4.7	1819	16 Q97K40	Q97K40 clostridium
63	133	4.7	711	16 Q99ZB3	Q99ZB3 streptococc
64	133	4.7	1313	16 Q9PQ16	Q9PQ16 ureaplasma
65	133	4.7	2500	5 Q96223	Q96223 plasmodium
66	132.5	4.6	427	10 Q03651	Q03651 hordeum vul
67	132.5	4.6	758	2 Q6007	Q6007 moraxella c
68	132.5	4.6	758	2 P94930	P94930 moraxella c
69	132.5	4.6	3029	16 Q55582	Q55582 synechocyst
70	132	4.6	485	2 Q53786	Q53786 streptococc
71	132	4.6	508	2 Q53995	Q53995 salmonella
72	132	4.6	566	2 Q32309	Q32309 bacillus th
73	131.5	4.6	1993	16 Q92F65	Q92F65 listeria in
74	131	4.6	498	16 Q97EM4	Q97EM4 clostridium
75	131	4.6	694	16 Q97EM5	Q97EM5 clostridium
76	131	4.6	3610	5 Q968T7	Q968T7 plasmodium
77	131	4.6	3620	5 Q968T6	Q968T6 plasmodium
78	130.5	4.6	439	16 Q97HA5	Q97HA5 clostridium
79	130.5	4.6	484	16 Q97Q49	Q97Q49 streptococc
80	130.5	4.6	583	2 Q45822	Q45822 clostridium
81	130.5	4.6	806	2 Q9X4F8	Q9X4F8 vibrio chol
82	130.5	4.6	2271	16 Q99QY4	Q99QY4 staphylococ
83	130.5	4.6	2364	2 Q46342	Q46342 clostridium
84	130	4.6	759	2 Q66006	Q66006 moraxella c
85	129.5	4.5	1166	2 Q86489	Q86489 staphylococ
86	129.5	4.5	1260	16 Q99ST8	Q99ST8 staphylococ
87	129.5	4.5	1849	2 Q9SAK2	Q9SAK2 lactobacill
88	129	4.5	826	10 Q9CAR6	Q9CAR6 arabidopsis
89	129	4.5	887	10 Q94A41	Q94A41 arabidopsis

90 129 4.5 1171 2 Q9KWX6 Q9KWX6 staphylococ
91 129 4.5 1349 16 Q92EV6 Q92EV6 listeria in
92 128 4.5 521 2 P71034 P71034 bacillus sp
93 128 4.5 521 2 P71034 P71034 bacillus sp
94 128 4.5 549 2 Q31193 Q31193 bacillus st
95 128 4.5 549 2 Q8WY6 Q8WY6 bacillus st
96 128 4.5 549 2 Q8WY6 Q8WY6 bacillus st
97 128 4.5 1212 5 Q900L0 Q900L0 plasmodium
98 128 4.5 1296 16 Q9P9W1 Q9P9W1 xyella fas
99 128 4.5 1477 2 Q480Z8 Q480Z8 haemophilus
100 128 4.5 1645 16 Q97MX7 Q97MX7 clostridium
101 128 4.5 1758 16 Q9TMS5 Q9TMS5 escherichia
102 128 4.5 1963 2 Q9XCQ3 Q9XCQ3 salmonella

ALIGNMENTS

RESULT 1
Q9LCJ7 PRELIMINARY; PRT: 1016 AA.
AC Q9LCJ7
DT 01-OCT-2000 (TREMELrel. 15, Created)
DT 01-OCT-2000 (TREMELrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMELrel. 19, Last annotation update)
DE DEXTRANSUCRASE.
GN DSRT.
OS Leuconostoc mesenteroides.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
OC Leuconostoc.
OX NCBI_TaxID=1245;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NRRL B-512F;
RX MEDLINE=20169623; PubMed=10705445;
RA Funane K., Mizuno K., Takahara H., Kobayashi M.;
RT "Gene encoding a dextranucrase-like protein in Leuconostoc
mesenteroides NRRL B-512F.";
RL Biosci. Biotechnol. Biochem. 64:29-38(2000).
DR EMBL; AB020020; BAA90527.1; -;
DR HSP; P06278; LVJS.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF02324; Glyco_hydro_70; 1.
SQ SEQUENCE 1016 AA; 110344 MW; 8996E8DE13CCCB47 CRC64;

Query Match 54.08; Score 1539; DB 2; Length 1016;
Best Local Similarity 54.98; Pred. No. 2.3e-72;
Matches 308; Conservative 70; Mismatches 151; Indels 32; Gaps 6;

QY 2 TYRGEFLANDIDNSNPVQAEOLNWLHYLLNFGTITANDQANFDSVRVYDAPDNIDAD 61
DB 467 TGGGYDFLLANDIDNSNPVQAEOLNWLHYLLNFGTITANDQANFDSVRVYDAPDNIDAD 526
QY 62 LMNTAQDYFNAAYGMD-SDAVSNKHINILEDNHNADPEYFNKIGNPOLTMDDTIKNSLNH 120
DB 527 LLOIADYFKAAYGVOKSDAISNQHVSILEDSWDAEYVVDKNGDGNLSMDNKLRLSLKY 586
QY 121 GLS-----DAFNKGLDAIVHQSADRENSTENYVIPNTSFVRAHDNNSQDQIQAIR 174
DB 587 SLTPFAVDQYGNKRSGLPELTNSLVDRDNTSDTONTAPNTSFVRAHDNSEVQTVIAELIK 646
QY 175 DVTGKDYHTE--TPDEKGDIDAYIQDNQSVKYNLKNIPASVAILTNKDTIPRVYVG 232
DB 647 QRIDPDSGLSPTMDQLFEAKFYNDQDKTKEQTQNPSTATILTNKDTIPRVYVG 706
QY 233 DLYTDGGQYMEHQTRYDPTLNLKSRVYKYGAGQSM-----QTMVGSGNNILTSV 284
DB 707 DMYTDGGYMATKSLYDAIDTLKSRVYKYGAGQSM-----QTMVGSGNNILTSV 766
QY 285 RYKGMATATDGTDETRQGGVGVVSTPRLKLVNDKYLHKAHKNQYRAVLTT 344
DB 767 RYKGMATATDGTDETRQGGVGVVSTPRLKLVNDKYLHKAHKNQYRAVLTT 826
QY 345 TDGVINYTSQDQAG--VAMTDENGULYLSHNLVNVNGKEADTAVQGVANPDVSGYLAV 401

DB 827 KDGITVYSDSDVSQSLIRYNSGQLLIFNSDIV-----GTANPOVSGYLAV 874
QY 402 WVPVGASDQDARTAPSTKNSGNSAYRTNAAFDSNVIFEAFSNFVYPTTKESERANVRI 461
DB 875 WVPVGASDQDARTESSTATTDDGOTLHNSAALDSQVIESFSNFQSTPTTEAEVANYQI 934
QY 462 AQNAOFFASLGTSTFEMAPQPNSSKDRFTFLDSTINGYAFTRDYDLGHSBPNKYGTDBDL 521
DB 935 ANNTDLYKSMGTNFEFPFQNSSTDSFLDSITQNGYAFTRDYDLGHSBPNKYGTDBDL 994
QY 522 RNAIOALHKAGLVNADWVVD 542
DB 995 RTAIRALHATGKAMADWVVD 1015

RESULT 2

Q9RE05 PRELIMINARY; PRT: 2057 AA.
AC Q9RE05
DT 01-MAY-2000 (TREMELrel. 13, Created)
DT 01-MAR-2001 (TREMELrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMELrel. 19, Last annotation update)
DE ALTERNANSUCRASE (EC 2.4.1.140).
GN ASR.
OS Leuconostoc mesenteroides.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
OC Leuconostoc.
OX NCBI_TaxID=1245;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NRRL B-1355;
RX MEDLINE=20080809; PubMed=10612736;
RA Arguello-Morales M.A.; Remaud-Simeon M., Pizut S., Sarcabal P.,
RT "Sequence analysis of the gene encoding alternansucrase, a sucrose
glucosyltransferase from Leuconostoc mesenteroides NRRL B-1355.";
RL FEMS Microbiol. Lett. 182:81-85(2000).
DR EMBL; AJ250173; CAB65910.2; -;
DR InterPro; IPR000604; Chlamydia_OMP.
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 12.
DR Pfam; PF02324; Glyco_hydro_70; 1.
DR ProDom; PD001717; Chlamydia_OMP; 1.
KW Transferase; Glycosyltransferase.
SQ SEQUENCE 2057 AA; 228987 MW; 62BCE9385D9A11BE CRC64;

Query Match 53.7%; Score 1532.5; DB 2; Length 2057;
Best Local Similarity 53.2%; Pred. No. 1.2e-71;
Matches 315; Conservative 85; Mismatches 137; Indels 55; Gaps 14;

QY 4 RGOEFLANDIDNSNPVQAEOLNWLHYLLNFGTITANDQANFDSVRVYDAPDNIDAD 63
DB 586 KGSEFLANDIDNSNPVQAEOLNWLHYLLNFGTITANDQANFDSVRVYDAPDNIDAD 645
QY 64 NIAQDYFNAAYGMD-SDAVSNKHINILEDNHNADPEYFNKIGNPOLTMDDTIK-----NSL 118
DB 646 KIADYFKALYGTGDSANANKHLSILEDNNGKDPQIVNQQGNAQLTMDYTVTSQFNSL 705
QY 119 NHGLSDATNRW-----GLDAIVHQSADREN-----NSTENVYIPNYSFVRA 160
DB 706 THGANNRSNMVYFLDTGYLNGDLNKKIVDKNRPNSGTLVNRANSQDTKVIPIVNFVRA 765
QY 161 HDNNSQDQIQAIRD--VTGKYDHTTFDEQKIDAYIQDN--STVKYNLYNPASY 216
DB 766 HDYDAQPLRKAMIDHGIKNMQDTFTFDQAQMEFYKQDENPNSGFKKYNLYNPASY 825
QY 217 AILLTNKDTIPRVYGDLYTDGGQYMEHQTRYDPTLNLKSRVYKYGAGQSMQTMVSG- 275
DB 826 AMLLTNKDTPRVYGDLYTDGGQYMEHQTRYDPTLNLKSRVYKYGAGQSMQTMVSG- 885

```

QY 276 ----GNNILTSVRYGKAMTATDTGTDEI-----RTQIGVVVSVNTPNLKLGVNDKVLH 327
Db 886 DLKDGFTDLTSVRRGKGMTSQTITQDMSQDYKKGIGVGVNPNPDLKLNNDKTKITLH 945
QY 328 MGAHKNQYRAAVALTTDGVINTSDQAPVAMTDENGDLXLSHNLVNGKEA---D 384
Db 946 MGRHKNQYRALVLSNDGIDYVSDDKAPLTNTDNGDLIFHKNTFV--KQDGTIIN 1003
QY 385 TAVQGYANPDVGYLAVVVPVGSADNODARTASTKNSGN--SAYRTNAAFDSNWIFEA 442
Db 1004 YEMKSLNALTISGYLVVVPVGSADNODARTASTKNSGN--SAYRTNAAFDSNWIFEA 1062
QY 443 FSNVYVPTKESBEANVRIQADFRASLGTSPFMAPOVNSK-----RTFLDSIDN 497
Db 1063 FSNFQAMPTSPQSTNVVVIATKANLFEKIGITISFELAPQYRSSGDTNYGSGFLDSPLNN 1122
QY 498 GYAFTRDYDLGMSR-----PNKYGTDEDLRNAIQALHKAGLQVMADWPDQ 543
Db 1123 GYAFTRDYDLGFKADGNPNPKYGTQDQLRNAIEALHKNQMAIADWPDQ 1174

RESULT 3
Q9ZAR4 PRELIMINARY; PRT; 1527 AA.
AC Q9ZAR4;
DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE DEXTRANSUCRASE.
GN DEX.
OS Leuconostoc mesenteroides.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
OC Leuconostoc.
OX NCBI_TaxID=1245;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL B-512-F;
RA Bhattacharya R., Singh D.K.S.;
RT "Cloning and Molecular Characterization of Dextranucrase Gene from
RL Leuconostoc mesenteroides NRRL B-512F."
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U81374; AAD10952.1; -.
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 16.
DR Pfam; PF02324; Glyco_hydro_70; 1.
SQ SEQUENCE 1527 AA; 169709 MW; 1DFAFA237C743398 CRC64;

```

```

Query Match 53.4%; Score 1523; DB 2; Length 1527;
Best Local Similarity 56.2%; Pred. No. 2.6e-71;
Matches 305; Conservative 68; Mismatches 150; Indels 20; Gaps 5;

QY 5 GQEFLLANDIDNSPPVQAEQLNWLILYLLNFGTITANNQANFDSVRYDAPDNIDALN 64
Db 503 GFELLANQEDNSVNVVEAEQLNWLILYLLNFGTITANDANFDSVRYDAPDNIDALN 562
QY 65 IAQDFNAAAGMD--SDAVSNKHINILEDNADPEYFNKIGNPQLTMDTITKNSLNHGLS 123
Db 563 IAADYFKLAYGVQDQANQHLSTLEDWSHNDPLVYTDGNSQLTMDYVHTQLNSLT 622
QY 124 DATN--RWGILDAIVHQSADRENNSSTENVIPNYSFVRAHDNNSQDIQNAIRDVTKGDYH 182
Db 623 KESDITRGTMQRVDYVWDRSDNSDSTENEAIPNYSFVRAHDSEVQVIAQYVSDLYPDYEN 682
QY 183 TP--TFEEQKIDAYIQDQSTVVKYNDLXNIPASAYALLTNKDTIPRVYGDLYTDGGQ 240
Db 683 SLAPTEQLAAAFKYVNEDEKLADKKYIQYNASAYAMLLTNKDTIPRVYGDLYTDGGQ 742
QY 241 YMEHOTRYDITLNLKSRVYAGQSMQTMVSGGNNILTSVRYGKAMTATDTGTDE 300
Db 743 YMATKSPYDAINTLTKARVQYVAGQSMVDS----NDVLTSVRYGKAMTATDTGTSE 798

```

```

QY 301 TRIQIGVVVSVNTPNLKLGVNDKVLHMGAAHKNQYRAAVALTTTDDGVINTSDQAPVA 360
Db 799 TRREGIGVIVSNNAELQLEDGHTVILHMGAAHKNQYRALLSTTADGLAYYTDENAPVA 858
QY 361 MTDENGDLYLSSHNLVNGKEADRTAVQGYANPDVSGYLAVVVPVGSADNODARTAPSTE 420
Db 859 YTDANGDLIFTNE-----SIYGVQNPQVSGYLAVVVPVGAQDQDARTASDTT 906
QY 421 KNSGNSAYRTNAAFDSNVIFEAFSNEVYPTKESBEANVRIQADFRASLGTSPFMAP 480
Db 907 TNSVDKVFHNSNAALDSQVIYEGFSNFQAFATSSSETNVVIAQNAQDFKQMGVTSFQAP 966
QY 481 QYSSSDRTFLDSTIDNGYAFIDRYDLGMSPEPNKYGTDEDLRNAIQALHKAGLQVMADW 540
Db 967 QYSSSDRTFLDSTIDNGYAFIDRYDLGMSPEPNKYGTDEDLRNAIQALHKAGLQVMADW 1026
QY 541 PDQ 543
Db 1027 PDQ 1029

RESULT 4
Q9EZHS PRELIMINARY; PRT; 1508 AA.
AC Q9EZHS;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE DEXTRANSUCRASE DSRB742.
GN DSRB742.
OS Leuconostoc mesenteroides.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
OC Leuconostoc.
OX NCBI_TaxID=1245;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B-742CB;
RA Kim H.-S., Kim D., Ryu H.-J., Robyt J.F.;
RT "Leuconostoc mesenteroides B-742CB, a dextranucrase gene."
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF294469; AAC38021.1; -.
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 14.
DR Pfam; PF02324; Glyco_hydro_70; 1.
SQ SEQUENCE 1508 AA; 168542 MW; E2FCFA0F87AB4F3A CRC64;

Query Match 51.5%; Score 1469.5; DB 2; Length 1508;
Best Local Similarity 53.7%; Pred. No. 1.6e-68;
Matches 294; Conservative 72; Mismatches 160; Indels 21; Gaps 6;

QY 1 QTYRGQEFLLANDIDNSPPVQAEQLNWLILYLLNFGTITANNQANFDSVRYDAPDNIDA 60
Db 481 QSKSGFELLANDIDNSPPVQAEQLNWLILYLLNFGTITANNQANFDSVRYDAPDNIDA 540
QY 61 DLMNIAQDYFNAAAGMD--SDAVSNKHINILEDNADPEYFNKIGNPQLTMDTITKNSLN 119
Db 541 DLLQIAADYFKAAAGVQDQANQHLSTLEDWSHNDPEYFNKIGNPQLTMDTITKNSLN 600
QY 120 HGLS--DATNREGLDAIVHQSADRENNSSTENVIPNYSFVRAHDNNSQDIQNAIRDV-- 176
Db 601 WSLTKDMRMGRMTQRFMDYLVNRRHDSSTENTAIPNYSFVRAHDSEVQVIAQIISLHP 660
QY 177 TGDYHTFTFDEDEKIDAYIQDQSTVVKYNDLXNIPASAYALLTNKDTIPRVYGDLYT 236
Db 661 DVKNLSAPTADQQLAEAFKIYNDEKQADKTYQTQYMPISAYAMLLTNKDTIPRVYGDLYT 720
QY 237 DGGVMEHOTRYDITLNLKSRVYAGQSMQTMVSGGNNILTSVRYGKAMTATDT 296
Db 721 DGGVMEHOTRYDITLNLKSRVYAGQSMQTMVSGGNNILTSVRYGKAMTATDT 776
QY 297 GTDETTRTQIGVIGVYVNTPNLKLGVNDKVLHMGAAHKNQYRAAVALTTTDDGVINTSDQ 356

```

```
Db 777 GNAATRTGIGIVSNKENLALKSGDTVTLLHGAHAKNQAFRLLLGTADNLSTYDND-N 835
QY 357 APVATMDENGDLXLSHNLVNGKEADTAVOGYANPDVSGYLAVWVPVPGASDNODART 416
Db 836 APVATMDENGDLXLSHNLVNGKEADTAVOGYANPDVSGYLAVWVPVPGASDNODART 883
QY 417 PSTENKNSGSAVRTNAAFDSNVIFEAFSNFYVTPTKESERANVRIAQNAOFFASLGFTSF 476
Db 884 SDDTAHDGKTFHNSAALDSQVIYEGFSNFQAFATNTEDYTNAVIAKNQGLFKDGIISF 943
QY 477 EMAPQVNSKORTFIDSTTDNGYAFTRDYLGMSPNKYGTDEDLRNAIQALHKGAGLQW 536
Db 944 QLAQYRSSTDTSFLLSIQNGYAFTRDYLGMSPNKYGTDEDLRNAIQALHKGAGLQW 1003
QY 537 ADWVPDQ 543
Db 1004 ADWVPDQ 1010

RESULT 5
Q9L466 PRELIMINARY; PRI: 1477 AA.
AC Q9L466;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE DEXTANSUCRASE (EC 2.4.1.5).
GN DSRG.
OS Leuconostoc mesenteroides.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
OC Leuconostoc.
OX NCBI_TaxID=1245;
RN [1]
RP SEQUENCE FROM M.A.
RC STRAIN=NRRL B-1355;
RA Arguello-Morales M.A., Remaud-Simeon M., Pizut S., Sarcabal P.,
RA Willenot R.M., Monsan P.;
RT "Sequence analysis of the gene encoding alternansucrase, a sucrose
RT glucosyltransferase from Leuconostoc mesenteroides NRRL B-1355."
RL Submitted (Oct-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ250172; CAB76565.1;
DR InterPro: IPR002479; CW_binding.
DR InterPro: IPR003318; Glyco_hydro_70.
DR Pfam: PF01473; CW_binding.1; 14.
DR Pfam: PF02324; Glyco_hydro_70; 1.
KW Transferase; Glycosyltransferase.
SQ SEQUENCE 1477 AA; 164887 MW; E6F5710D8DFCB831 CRC64;

Query Match 51.5%; Score 1467.5; DB 2; Length 1477;
Best Local Similarity 53.7%; Pred. No. 1.9e-68;
Matches 294; Conservative 71; Mismatches 161; Indels 21; Gaps 6;

QY 1 QTYRGQEFLLANDIDNSNPVQAEQLNWLHNFQITANNQDANFSDVRYDAPDNIDA 60
Db 450 QSKGGFELLANDVDSNPVQAEQLNWLHNFQITANNQDANFSDVRYDAPDNIDA 509
QY 61 DLNIAQDYFNAAYGM-SDAVSNKHINILEDNNHADPEYFNKGNPQLTMDTINKSLN 119
Db 510 DLLQIAADYFNAAYGM-SDAVSNKHINILEDNNHADPEYFNKGNPQLTMDTINKSLN 569
QY 120 HGLSDATN-RWGLDAIVHQSLADRENSTENVIPNYSFVRAHDNNSQDQIONAIRV-- 176
Db 570 WSLTMDMRGTMRQFMDYLYLNRNHSSTENTAIPTNYSFVRAHDNNSQDQIONAIRV-- 629
QY 177 TGDYHTTFEEDQKIDAYIQDQNSTVKYKYNLNPASVAILTNKDTIPRYVGLYT 236
Db 630 DVKNSLAPADQLAEAFKIYNNDKQADKKYQTNMPSAYAMLLTNKDTIPRYVGLYT 689
QY 237 DGGQYWFQYTDYTLNLLSRVYKIVAGGSMQTSYGGNNNLTISVRYKGGKATATDT 296
Db 690 DGGQYMANSPYFDALNGLLAKRIKIVAGGSMQTSYGGNNNLTISVRYKGGKATATDT 745
```

```
QY 297 GTDETRTQIGIVSNKPNLKLGVNDXVVLHMGAAHKNQYRAAVLTTTGVINYSDDQ 356
Db 746 GNAATRTGIGIVSNKENLALKSGDTVTLLHGAHAKNQAFRLLLGTADNLSTYDND-N 804
QY 357 APVATMDENGDLXLSHNLVNGKEADTAVOGYANPDVSGYLAVWVPVPGASDNODART 416
Db 805 APVATMDENGDLXLSHNLVNGKEADTAVOGYANPDVSGYLAVWVPVPGASDNODART 852
QY 417 PSTENKNSGSAVRTNAAFDSNVIFEAFSNFYVTPTKESERANVRIAQNAOFFASLGFTSF 476
Db 853 SDDTAHDGKTFHNSAALDSQVIYEGFSNFQAFATNTEDYTNAVIAKNQGLFKDGIISF 912
QY 477 EMAPQVNSKORTFIDSTTDNGYAFTRDYLGMSPNKYGTDEDLRNAIQALHKGAGLQW 536
Db 913 QLAQYRSSTDTSFLLSIQNGYAFTRDYLGMSPNKYGTDEDLRNAIQALHKGAGLQW 972
QY 537 ADWVPDQ 543
Db 973 ADWVPDQ 979

RESULT 6
O69388 PRELIMINARY; PRI: 1455 AA.
AC O69388;
DT 01-AUG-1998 (Tremblrel. 07, Created)
DT 01-AUG-1998 (Tremblrel. 07, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE GLUCOSYLTRANSFERASE-SI.
GN GTFG.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM M.A.
RC STRAIN=MT4239;
RX MEDLINE=98231643; Pubmed=9570124;
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,
RA Kimura S., Hamada S.;
RT "Molecular analyses of glucosyltransferase genes among strains of
RT Streptococcus mutans."
RL FEMS Microbiol. Lett. 161:331-336(1998).
DR EMBL: D88658; BAA26110.1;
DR InterPro: IPR002479; CW_binding.
DR InterPro: IPR003318; Glyco_hydro_70.
DR Pfam: PF01473; CW_binding.1; 10.
DR Pfam: PF02324; Glyco_hydro_70; 1.
KW Transferase.
SQ SEQUENCE 1455 AA; 163046 MW; 6D90A4978D35DD82 CRC64;

Query Match 51.4%; Score 1466.5; DB 2; Length 1455;
Best Local Similarity 52.9%; Pred. No. 2.1e-68;
Matches 292; Conservative 94; Mismatches 141; Indels 25; Gaps 8;

QY 1 QTYRGQEFLLANDIDNSNPVQAEQLNWLHNFQITANNQDANFSDVRYDAPDNIDA 60
Db 425 RTGGYEFLLANDVDSNPVQAEQLNWLHNFQITANNQDANFSDVRYDAPDNIDA 484
QY 61 DLNIAQDYFNAAYGM-SDAVSNKHINILEDNNHADPEYFNKGNPQLTMDTINKSLN 119
Db 485 DLLQIAADYFNAAYGM-SDAVSNKHINILEDNNHADPEYFNKGNPQLTMDTINKSLN 544
QY 120 HGLSDATN-RWGLDAIVHQSLADRENSTENVIPNYSFVRAHDNNSQDQIONAIRV-- 174
Db 545 YSLAKPLNQSGMNPITLNSLVNKTDDNAETAAPVPSYFVRAHDSVQDLIRNIRAEIN 604
QY 175 -DVTGKYHTTFEEDQKIDAYIQDQNSTVKYKYNLNPASVAILTNKDTIPRYVGL 233
Db 605 PNYYG---YSFTMEIKKAFETIYNKOLLATKEKTYHTNTALSYALLITNKSSPVRYGD 661
```

QY 234 LYTGGQYMEHOTRYDYTLNLLKSRVYVAGSQSMQTSVGGNNNLTSTRYKGGAMTA 293
Db 662 MFTDDGQYMAHKTINTEAETLLKARIKYVSGQAMRNQOV - GNSLIITSVRVKGALKA 720
QY 294 TDGTGDETGTQIGVGVVSTPNLKLGVNDKVLHMGAAHKNQOYRAAVLTITTDGVTNYTS 353
Db 721 TDGTGDETGTQIGVGVVSTPNLKLGVNDKVLHMGAAHKNQOYRAAVLTITTDGVTNYTS 780
QY 354 DQGAP--VAMTDENGDLYLSSHNLVYNGKEADTAVQGYANPDVSGYLAWVPVPGASDNQ 411
Db 781 DOEAAGLVRYTNDGELIETA-----AD--IKGYANPQVSGYLGWVVPVPGAAADQ 828
QY 412 DARTAPSTKNSGNSAYKTNAAFDNSNVIFEARNFVYTPPKESERANVRITAOADFFASL 471
Db 829 DVRTAAPTSTDGKSVHQMAALDSRVMEGEFNFQAFATKEEYTVVIAKNVDFAEW 888
QY 472 GFTSFEMAPYNSKDRFTLDSTIDNGYAFTRDYDGLGMSPEPNKYGTDEDLRNAIQALHKA 531
Db 889 GVTDFEMAPYNSKDRFTLDSTIDNGYAFTRDYDGLGMSPEPNKYGTDEDLRNAIQALHKA 948
QY 532 GLOVADWVDPQ 543
Db 949 GIKVADWVDPQ 960
RESULT 7
OS2224 PRELIMINARY; PRT; 1508 AA.
AC OS2224;
DT 01-JUN-1998 (T-EMBLrel. 06, Created)
DT 01-JUN-1998 (T-EMBLrel. 06, Last sequence update)
DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)
DE GLUCOSYLTRANSFERASE (EC 2.4.1.5).
GN DSRB.
OS Leuconostoc mesenteroides.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
OC Leuconostoc.
OX NCBI_TaxID=1245;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NRRL B-1299;
RA Monchois V., Renaud-Simeon M., Monsan P., Willemot R.M.;
RT "Cloning and sequencing of a gene coding for an extracellular
RT dehydrogenase (DSRB) from Leuconostoc mesenteroides NRRL B-1299
RT synthesizing only a a(1-6) glucan."
RL FEMS Microbiol. Lett. 0:0-0(1998).
DR EMBL; AF030129; AAB95453.1; -
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 14.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Transferase; Glycosyltransferase.
SQ SEQUENCE 1508 AA; 168511 MW; E70CECB57A70D1F0 CRC64;

Query Match 51.3%; Score 1464.5; DB 2; Length 1508;
Best Local Similarity 53.6%; Pred. No. 2.8e-68;
Matches 293; Conservative 72; Mismatches 161; Indels 21; Gaps 6;

QY 1 QTYRGQEFLLANDIDNSNPVVOAEQLNWLHYLLNFGPITANNDQANFDSVRVDPADNIDA 60
Db 481 QSKGFEELLANDVNSNPVVOAEQLNWLHYLLNFGPITANNDQANFDSVRVDPADNIDA 540
QY 61 DLMNTAQDYFNAAYGM--SDAVSNKHINILEDWNHADPEYFNKIGNPQLTMDDTIKNSLN 119
Db 541 DLLQTAADYFKAAYGVDRKNDATANOHLISILEDWSDNDPEYVKKDFGNQLTMDDTIKNSLN 600
QY 120 HGLSDATN-RWGLDAIVHQSLADRENSTENVIPNYSFVRAHNNNSQDQONARV-- 176
Db 601 WSLYKDMRMGTQMRFDYLYLNRNHDSTENTATPNYSFVRAHNSQDQONARV-- 660
QY 177 TGDYHTTTFDEQKIDAYIQDQNSTVKKYNLXNIPASYAILLTKNDTIPRVYGDLYT 236
Db 177 TGDYHTTTFDEQKIDAYIQDQNSTVKKYNLXNIPASYAILLTKNDTIPRVYGDLYT 236

Db 661 DVKNSLAPTADQLAEAFKIYNNDKQADKKYQYNNPISAYAMLLTNKDTVPRVYGLYT 720
QY 237 DGGQYMEHOTRYDYTLNLLKSRVYVAGSQSMQTSVGGNNNLTSTRYKGGAMTADT 296
Db 721 DGGQYMEHOTRYDYTLNLLKSRVYVAGSQSMQTSVGGNNNLTSTRYKGGAMTADT 776
QY 297 GTDERTQIGVGVVSTPNLKLGVNDKVLHMGAAHKNQOYRAAVLTITTDGVTNYTSQOG 356
Db 777 GTDERTQIGVGVVSTPNLKLGVNDKVLHMGAAHKNQOYRAAVLTITTDGVTNYTSQOG 835
QY 357 APVAMTDENGDLYLSSHNLVYNGKEADTAVQGYANPDVSGYLAWVPVPGASDNODARTA 416
Db 836 APVAMTDENGDLYLSSHNLVYNGKEADTAVQGYANPDVSGYLAWVPVPGASDNODARTA 883
QY 417 PSTKNSGNSAYKTNAAFDNSNVIFEARNFVYTPPKESERANVRITAOADFFASLGFSTF 476
Db 884 SDDTAHOGKTHSNALDSQVYIEGFSNFQAFATNTEDITNAVIAKNQLEKDGITSF 943
QY 477 EMAPQYNSKDRFTLDSTIDNGYAFTRDYDGLGMSPEPNKYGTDEDLRNAIQALHKAQVLM 536
Db 944 QLAPOYRESSTDTSLDSIIQNGYAFTRDYDGLGVTGTPKYGTVDLRLDAIKALHANGIQAI 1003
QY 537 ADMVDPQ 543
Db 1004 ADMVDPQ 1010
RESULT 8
O69397 PRELIMINARY; PRT; 1455 AA.
ID AC O69397;
DT 01-AUG-1998 (T-EMBLrel. 07, Created)
DT 01-AUG-1998 (T-EMBLrel. 07, Last sequence update)
DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)
DE GLUCOSYLTRANSFERASE-Si.
GN GTFC.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MT4467;
RX MEDLINE-98231643; PubMed-9570124;
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,
RA Kimura S., Hamada S.;
RT "Molecular analyses of glucosyltransferase genes among strains of
RT Streptococcus mutans."
RL FEMS Microbiol. Lett. 161:331-336(1998).
DR EMBL; D89978; BAA36120.1; -
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 9.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Transferase.
SQ SEQUENCE 1455 AA; 162914 MW; A1263427BF24E8E1 CRC64;

Query Match 51.1%; Score 1458.5; DB 2; Length 1455;
Best Local Similarity 52.5%; Pred. No. 5.6e-68;
Matches 290; Conservative 96; Mismatches 141; Indels 25; Gaps 8;

QY 1 QTYRGQEFLLANDIDNSNPVVOAEQLNWLHYLLNFGPITANNDQANFDSVRVDPADNIDA 60
Db 425 RTGGYEFLLANDVNSNPVVOAEQLNWLHYLLNFGPITANNDQANFDSVRVDPADNIDA 484
QY 61 DLMNTAQDYFNAAYGM--SDAVSNKHINILEDWNHADPEYFNKIGNPQLTMDDTIKNSLN 119
Db 485 DLLQIAGDYLKAAGKGIHKNKANDHLISILEAWSYNDTPYLHDDGDNMINNDNRLSL 544
QY 120 HGLSDATN-RWGLDAIVHQSLADRENSTENVIPNYSFVRAHNNNSQDQONARV-- 174
Db 545 YSLAKPLNQRSGNNPLTNSLVNRTDDNAETAAPVPSYFIRAHDSQVLDLTIKAIKIN 604


```
QY 175 -DVTGKDYHTTFEDBQKIDAYIQDNSTVKYKYNLYNPASVAILLTKOTIPRVYGD 233
DB 605 PNVVG---YSFTMEELKAFIYNKOLLATERKYYTHYNTALSYALLTNKSSVPRVYGD 661
QY 234 LYTDGGQYMEHQTRYDITLNLKSKRVKVVAGQSMQTSVGGNNILTSVRYGKAMTA 293
DB 662 MFTDDGQYMAHKTINYEALTELLKARIKIVSGQAMRNQOV--GNSIITSVRYGKALKA 720
QY 294 TDGTDETRTQIGVYVSVNTPNLKGVNDKVLHMGAAHKNOQYRAAVLTDTDGVYNTS 353
DB 721 TDGTDETRTQIGVYVSVNTPNLKGVNDKVLHMGAAHKNOQYRAAVLTDTDGVYNTS 780
QY 354 DQGAP--VAMTDENSLIYSSHNLVNGKEADTAVOGYANPDVSGYLAVWVPVGSADNQ 411
DB 781 DQEAAGLVRYTNDRGELIFTA-----AD--IKGVANPOVSGYLGWVPVGAADQ 828
QY 412 DARTAPSTEKNSGNSAYRTNAAFDSNVIFEAFSNFYVYPTTKESERANVRIAQNAFASL 471
DB 829 DVAARAAPSTDGKSVHQNAALDSRVNPEGSFNEQAFATKKEEYTNVVIKAVDKFAEW 888
QY 472 GFTSFEMAPQYNSKDRFTDSTIDNGYAFTRDYLGMSEPNKYGTDEDLRNAIQALHKA 531
DB 889 GVTDFEMAPQYVSSDGSFLDSVIQNGYAFTRDYLGIKPNKYGTADDLYKAIKALHKS 948
QY 532 GLQVMADWVPDQ 543
DB 949 GIKVMADWVPDQ 960

RESULT 9
O69382 PRELIMINARY; PRT: 1455 AA.
AC O69382;
DT 01-AUG-1998 (T-REMBLrel. 07, Created)
DT 01-AUG-1998 (T-REMBLrel. 07, Last sequence update)
DT 01-DEC-2001 (T-REMBLrel. 19, Last annotation update)
DE GLUCOSYLTRANSFERASE-SI.
GN GTFC.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MT8148;
RX MEDLINE=98231643; PubMed=9570124;
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,
RA Kimura S., Hamada S.;
RT "Molecular analyses of glucosyltransferase genes among strains of
RT Streptococcus mutans.";
RL FEMS Microbiol. Lett. 161:331-336(1998).
DR EMBL; D88652; BAA26102.1;
DR InterPro; IPR002479; CW_binding.
DR Pfam; PF01473; CW_binding_1; 9.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Transferase.
SQ SEQUENCE 1455 AA; 162970 MW; 27D4D3A1BECA2939 CRC64;

Query Match 51.0%; Score 1453.5; DB 2; Length 1455;
Best Local Similarity 52.8%; Pred. No. 1e-67;
Matches 290; Conservative 96; Mismatches 141; Indels 25; Gaps 8;

QY 1 QTYRGQEFLLANDIDNSPNVQAQLNWLFLNFGITANNDQANFDSVRVDAPDNIDA 60
DB 425 RTIGGYEFLLANDVDSNPNVQAQLNWLFLNFGIYANDPANFDSIRVDVNDVA 484
QY 61 DLMIATQDYFNAAYGM--DSDAVSNKHINILEDNHNADPEYFNKIGNPQLTMDTITKNSLN 119
DB 485 DLIQTAGDYLAANKGIHKNDKANDHLSILEANSYNQPTYLHDDGQWNNMDNRLSL 544
```

```
QY 120 HGLSDATN-RWGLDAIVHQSLADRENNSTENVIPNYSFVRAHDNNSQDQIONAIR---- 174
DB 545 YSLAKPLNQRSGMNFLLITNSLVNRTDDNAETAAPFSYSFIRAHSEVQDLIRDIKABIN 604
QY 175 -DVTGKDYHTTFEDBQKIDAYIQDNSTVKYKYNLYNPASVAILLTKOTIPRVYGD 233
DB 605 PNVVG---YSFTMEELKAFIYNKOLLATERKYYTHYNTALSYALLTNKSSVPRVYGD 661
QY 234 LYTDGGQYMEHQTRYDITLNLKSKRVKVVAGQSMQTSVGGNNILTSVRYGKAMTA 293
DB 662 MFTDDGQYMAHKTINYEALTELLKARIKIVSGQAMRNQOV--GNSIITSVRYGKALKA 720
QY 294 TDGTDETRTQIGVYVSVNTPNLKGVNDKVLHMGAAHKNOQYRAAVLTDTDGVYNTS 353
DB 721 TDGTDETRTQIGVYVSVNTPNLKGVNDKVLHMGAAHKNOQYRAAVLTDTDGVYNTS 780
QY 354 DQGAP--VAMTDENSLIYSSHNLVNGKEADTAVOGYANPDVSGYLAVWVPVGSADNQ 411
DB 781 DQEAAGLVRYTNDRGELIFTA-----AD--IKGVANPOVSGYLGWVPVGAADQ 828
QY 412 DARTAPSTEKNSGNSAYRTNAAFDSNVIFEAFSNFYVYPTTKESERANVRIAQNAFASL 471
DB 829 DVAARAAPSTDGKSVHQNAALDSRVNPEGSFNEQAFATKKEEYTNVVIKAVDKFAEW 888
QY 472 GFTSFEMAPQYNSKDRFTDSTIDNGYAFTRDYLGMSEPNKYGTDEDLRNAIQALHKA 531
DB 889 GVTDFEMAPQYVSSDGSFLDSVIQNGYAFTRDYLGIKPNKYGTADDLYKAIKALHKS 948
QY 532 GLQVMADWVPDQ 543
DB 949 GIKVMADWVPDQ 960

RESULT 10
O69385 PRELIMINARY; PRT: 1390 AA.
AC O69385;
DT 01-AUG-1998 (T-REMBLrel. 07, Created)
DT 01-AUG-1998 (T-REMBLrel. 07, Last sequence update)
DT 01-JUN-2001 (T-REMBLrel. 17, Last annotation update)
DE GLUCOSYLTRANSFERASE-SI.
GN GTFC.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MT4245;
RX MEDLINE=98231643; PubMed=9570124;
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,
RA Kimura S., Hamada S.;
RT "Molecular analyses of glucosyltransferase genes among strains of
RT Streptococcus mutans.";
RL FEMS Microbiol. Lett. 161:331-336(1998).
DR EMBL; D88655; BAA26106.1;
DR InterPro; IPR002479; CW_binding.
DR Pfam; PF01473; CW_binding_1; 7.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Transferase.
SQ SEQUENCE 1390 AA; 155375 MW; 8847E4956EF05E9F CRC64;

Query Match 50.9%; Score 1452.5; DB 2; Length 1390;
Best Local Similarity 52.5%; Pred. No. 1.1e-67;
Matches 290; Conservative 94; Mismatches 143; Indels 25; Gaps 8;

QY 1 QTYRGQEFLLANDIDNSPNVQAQLNWLFLNFGITANNDQANFDSVRVDAPDNIDA 60
DB 425 RTIGGYEFLLANDVDSNPNVQAQLNWLFLNFGIYANDPANFDSIRVDVNDVA 484
QY 61 DLMIATQDYFNAAYGM--DSDAVSNKHINILEDNHNADPEYFNKIGNPQLTMDTITKNSLN 119
```

```

Db 485 DLLOIAGDYLAAGAIIHKNDKAANDHLSLEAMSDNDTPYLHDDGDNMNMNKNLRLSLL 544
QY 120 HGLSDATN-RWGLDAIVHQSLADRENNSTENVIPNYSFVRADHNSQDQIONAIR-----174
Db 545 FSLAKPLNQRSGMPLTNSLVNRTDDNAETAAPVPSFIRAHDSVQDLRIIRAEIN 604
QY 175 -DVTGDKYHTTFDEDEQKIDAYIQDQNSTVKYKYNLNPASVAILLTNKTDTIPRVYIGD 233
Db 605 PNWVG---YSEFMEEIKKAFIYNKOLLATEKTYTHNTALSYALLTNKSSVPRVYIGD 661
QY 234 LYTDGGOYMEHOTRYDTLNLKSRVYVAGGQSMQTSVGGNNILTSVRVYKGGAMTA 293
Db 662 MFTDGGQYMAHKTINYEAIETLLAKRIYVGGQAMRNQV-GNSEIITSVRVYKGGALKA 720
QY 294 TDGTDETRTGGIGVYVNTPNLKLGVNDKVLHMGAAHKNQYRAAVLTITDGVINYTS 353
Db 721 TDGTDRTRTSGVAVIEGNNPSLRKASDRVVYVNMGAHKNQYRPLLTITDNGIKAYHS 780
QY 354 DOGAP--VAMTDENGDIYLSHNLVNGKEADTAVUGYANPDVSGYLAVWVPVGAADNQ 411
Db 781 DQEAAGLVRYTNDRGELIFTA-----AD--IKGYANPQVSGYLGWVPVGAADQ 828
QY 412 DARTAPSTEKNSGNSAYRTNAFDSNVIFAFSNFYVTPPKESERANVRIAQNADEPASIL 471
Db 829 DRYVAASTAPSTDGSKSVHQNALDSRVAFEGFSNFQAFATKKEBYTNVIAKNVYKFAEW 888
QY 472 GFTSFEMAPQVNSKDRFTLDSTIDNGYAFTRDYDLGMSBPNKYGTDEDLRNATQALHKA 531
Db 889 GVIDFEMAPQVNSSTGDSFLDSVQNGYAFTRDYDLGISKPNKYGTADOLVKAIALKHS 948
QY 532 GLQYMAWVDPQ 543
Db 949 GIKYMAWVDPQ 960

RESULT 11
ID O69391 PRELIMINARY; PRT; 1455 AA.
AC O69391;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE GLUCOSYLTRANSFERASE-SI.
GN GTF.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MT4251;
RX MEDLINE=98231643; PubMed=9570124;
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,
RA Kimura S., Hamada S.;
RT "Molecular analyses of glucosyltransferase genes among strains of
RT Streptococcus mutans". Lett. 161:331-336(1998).
RL FEMS Microbiol. Lett. 161:331-336(1998).
DR EMBL; D88661; BAA26114.1; -
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 9.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW transferase.
SQ SEQUENCE 1455 AA; 162804 MW; 683A359D073E9E1A CRC64;

Query Match 50.9%; Score 1451.5; DB 2; Length 1455;
Best Local Similarity 52.68%; Pred. No. 1.3e-67;
Matches 290; Conservative 93; Mismatches 143; Indels 25; Gaps 8;

QY 2 TYRGOEFLANDINSNPVVQAEQLNWLHFLNFGNIYANDPDANFDSIRVDAYDNVDAD 485

```

```

Db 426 TIGGYEFLANDVDSNPNVQAEQLNWLHFLNFGNIYANDPDANFDSIRVDAYDNVDAD 485
QY 62 LMNIAQDYFNAAYGM-DSDAVSNKHINILEDNHADPEYFNKIGNPQLTMDDTIKNSLH 120
Db 486 LLQIAGDYLAAGAIIHKNDKAANDHLSLEAMSDNDTPYLHDDGDNMNMNKNLRLSLL 545
QY 121 GLSDATN-RWGLDAIVHQSLADRENNSTENVIPNYSFVRADHNSQDQIONAIR-----174
Db 546 SLAKPLNQRSGMPLTNSLVNRTDDNAETAAPVPSFIRAHDSVQDLRIIRAEINP 605
QY 175 DVTGDKYHTTFDEDEQKIDAYIQDQNSTVKYKYNLNPASVAILLTNKTDTIPRVYIGD 234
Db 606 NVVG---YSEFMEEIKKAFIYNKOLLATEKTYTHNTALSYALLTNKSSVPRVYIGD 662
QY 235 YTDGGOYMEHOTRYDTLNLKSRVYVAGGQSMQTSVGGNNILTSVRVYKGGAMTA 294
Db 663 FTUDGGOYMAHKTINYEAIETLLAKRIYVGGQAMRNQV-GNSEIITSVRVYKGGALKA 721
QY 295 DTGDETRTGGIGVYVNTPNLKLGVNDKVLHMGAAHKNQYRAAVLTITDGVINYTS 354
Db 722 DTGDRTRTSGVAVIEGNNPSLRKASDRVVYVNMGAHKNQYRPLLTITDNGIKAYHS 781
QY 355 QGAP--VAMTDENGDIYLSHNLVNGKEADTAVUGYANPDVSGYLAVWVPVGAADNQ 412
Db 782 QEAAGLVRYTNDRGELIFTA-----AD--IKGYANPQVSGYLGWVPVGAADQ 829
QY 413 ARTAPSTEKNSGNSAYRTNAFDSNVIFAFSNFYVTPPKESERANVRIAQNADEPASIL 472
Db 830 VRYAASTAPSTDGSKSVHQNALDSRVAFEGFSNFQAFATKKEBYTNVIAKNVYKFAEW 889
QY 473 FTSFEMAPQVNSKDRFTLDSTIDNGYAFTRDYDLGMSBPNKYGTDEDLRNATQALHKA 532
Db 890 VTFEMAPQVNSSTGDSFLDSVQNGYAFTRDYDLGISKPNKYGTADOLVKAIALKHS 949
QY 533 LQYMAWVDPQ 543
Db 950 IKYMAWVDPQ 960

RESULT 12
Q59983
ID Q59983 PRELIMINARY; PRT; 1590 AA.
AC Q59983;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE GLUCOSYLTRANSFERASE-I PRECURSOR (EC 2.4.1.5).
GN GTFI.
OS Streptococcus sobrinus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1310;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OM2176;
RX MEDLINE=94146405; PubMed=8312602;
RA Sato S., Inoue M., Handa N., Aizawa Y., Isobe Y., Katayama T.;
RT "DNA sequence of the glucosyltransferase gene of serotype d
RT Streptococcus sobrinus".
RL DNA Seq. 4:19-27(1993).
DR EMBL; D13858; BAA02976.1; -
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 16.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Signal; Transferase; Glycosyltransferase.
FT SIGNAL 1 38 POTENTIAL.
FT CHAIN 39 1590 GLUCOSYLTRANSFERASE-I.
SQ SEQUENCE 1590 AA; 175955 MW; C3C83A57CF3C2B0E CRC64;

Query Match 49.6%; Score 1414; DB 2; Length 1590;
Best Local Similarity 52.7%; Pred. No. 1.3e-65;

```

Matches 289; Conservative 84; Mismatches 151; Indels 24; Gaps 9;

QY 5 GQFLLANDIDNSNPVQAEQLNWLHLYLLNFCTITANNQANFDSVRVADPONDIDLNM 64
DB 399 GYFLLANDVDNSNPVQAEQLNWLHLYLLNFCTITANNQANFDSVRVADPONDIDLNM 458
QY 65 IADYFNAAYGMD-SDAVSNKHINLEEDWNHADPEYFNKIGNPQLTMDDTIKNSLNHGLS 123
DB 459 ISSDYLKAAAYGIDKNNKANNHYSIWEASNDTPLYLHDDGDNLMNDKFRLSMLWSLA 518
QY 124 DATN-RWGLDAIVHQSADRENSTENVIPNVSFVRAHNNSDQIQONAIR-DVTGKDY 181
DB 519 KPLDKRGLAPLHNSLVDRVDDREVEYVPSYFARADHSEVQDIIRDLIKAEINPNSF 578
QY 182 -HTFTFDEQKIDAYIQDQNSTVKKYNLYNIPASVAILLTNKTIPRVYIGDLYTDGQ 240
DB 579 GYSFTQEEIQAFKIYNEDLKKTKKYYHNVPLSYTLTLLTNKGSIPRVYIGDMFTDDGQ 638
QY 241 YMEHQTRYDYLTLNLLKSRVYVAGGQSMQTSVGGNNILTSVRYGKGAMTATDGTDE 300
DB 639 YMAKTYNYDAIESLLKARKMYVSGGQAMQYQI-GNGEILT SVRYGKGALQSKGDAT 697
QY 301 TRTGIGVYVSNTPNLKLGVDKVV-LHMGAAHKNQYRAAVLTDTGVTNYTSDQGAP- 358
DB 698 TRTSGVGVVGNQPNFSL--DGKVALNMGAHAHQEYRALMSTKDGATYATDADASK 755
QY 359 ---VAMTDENGDIYLSHNLVNGKEADTAVQGYANPDVSGYLVAVVPGASNDQDART 415
DB 756 AGLVKRTDENGILYFLNDDL-----KGVANPQVSGFLQVWVPVGAADQDIRV 803
QY 416 APSTEKNSGNSAYETNAAFDSNVIFEAFFSVYPTPKESERANVRIAQNADFASLGFTS 475
DB 804 AASTASTDGKSLHQAADMSRVNFEFSNFSQFATKEEYTNVIANNVKPVSWGITD 863
QY 476 FEMAPQYNSKDRFLDSTIDNGYAFTRDYDLGMBSPNKYGTDEDLRNAIQALHAKGLQV 535
DB 864 FEMAPQYVSTDDGFLDSVQNGYAFTRDYDLGMBSPNKYGTADQLVKAIKALHAKGLV 923
QY 536 MADWVPDQ 543
DB 924 MADWVPDQ 931

RESULT 13
Q55263 PRELIMINARY; PRT; 1590 AA.
AC Q55263;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE GTF-I.
GN GLUCOSYLTRANSFERASE.
OS Streptococcus sobrius.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1310;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33478;
RA Sato S.;
RT "DNA and amino-acid sequences of water-insoluble-glucan synthetase
produced from Streptococcus sobrius ATCC 33478.";
RL Ann. Kagoshima Univ. Dental School 16:23-29(1996).
DR EMBL; D63570; BAA09792.1; -.
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 15.
DR Pfam; PF02324; Glyco_hydro_70; 1.
SQ SEQUENCE 1590 AA; 176058 MW; 9DF7A3F2C6E4FD43 CRC64;

Query Match 49.4%; Score 1409; DB 2; Length 1590;
Best Local Similarity 52.7%; Pred. No. 2.4e-65;

Matches 289; Conservative 82; Mismatches 153; Indels 24; Gaps 9;

QY 5 GQFLLANDIDNSNPVQAEQLNWLHLYLLNFCTITANNQANFDSVRVADPONDIDLNM 64
DB 399 GYFLLANDVDNSNPVQAEQLNWLHLYLLNFCTITANNQANFDSVRVADPONDIDLNM 458
QY 65 IADYFNAAYGMD-SDAVSNKHINLEEDWNHADPEYFNKIGNPQLTMDDTIKNSLNHGLS 123
DB 459 ISSDYLKAAAYGIDKNNKANNHYSIWEASNDTPLYLHDDGDNLMNDKFRLSMLWSLA 518
QY 124 DATN-RWGLDAIVHQSADRENSTENVIPNVSFVRAHNNSDQIQONAIR-DVTGKDY 181
DB 519 KPLDKRGLAPLHNSLVDRVDDREVEYVPSYFARADHSEVQDIIRDLIKAEINPNSF 578
QY 182 -HTFTFDEQKIDAYIQDQNSTVKKYNLYNIPASVAILLTNKTIPRVYIGDLYTDGQ 240
DB 579 GYSFTQEEIQAFKIYNEDLKKTKKYYHNVPLSYTLTLLTNKGSIPRVYIGDMFTDDGQ 638
QY 241 YMEHQTRYDYLTLNLLKSRVYVAGGQSMQTSVGGNNILTSVRYGKGAMTATDGTDE 300
DB 639 YMAKTYNYDAIESLLKARKMYVSGGQAMQYQI-GNGEILT SVRYGKGALQSKGDAT 697
QY 301 TRTGIGVYVSNTPNLKLGVDKVV-LHMGAAHKNQYRAAVLTDTGVTNYTSDQGAP- 358
DB 698 TRTSGVGVVGNQPNFSL--DGKVALNMGAHAHQEYRALMSTKDGATYATDADASK 755
QY 359 ---VAMTDENGDIYLSHNLVNGKEADTAVQGYANPDVSGYLVAVVPGASNDQDART 415
DB 756 AGLVKRTDENGILYFLNDDL-----KGVANPQVSGFLQVWVPVGAADQDIRV 803
QY 416 APSTEKNSGNSAYETNAAFDSNVIFEAFFSVYPTPKESERANVRIAQNADFASLGFTS 475
DB 804 AASTASTDGKSLHQAADMSRVNFEFSNFSQFATKEEYTNVIANNVKPVSWGITD 863
QY 476 FEMAPQYNSKDRFLDSTIDNGYAFTRDYDLGMBSPNKYGTDEDLRNAIQALHAKGLQV 535
DB 864 FEMAPQYVSTDDGFLDSVQNGYAFTRDYDLGMBSPNKYGTADQLVKAIKALHAKGLV 923
QY 536 MADWVPDQ 543
DB 924 MADWVPDQ 931

RESULT 14
Q68542 PRELIMINARY; PRT; 1449 AA.
AC Q68542;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE GLUCOSYLTRANSFERASE N (FRAGMENT).
GN GTFN.
OS Streptococcus salivarius.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1304;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=V1477;
RA Jaffe R.I.;
RT "Streptococcus salivarius V1477 gtfN";
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF049609; AAC05156.1; -.
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 8.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW transferase.
FT NON_TER 1449 1449
SQ SEQUENCE 1449 AA; 159895 MW; 0700F6D748471BFB CRC64;

Query Match 49.4%; Score 1407.5; DB 2; Length 1449;

SQ	SEQUENCE	1541 AA; 169370 MW; CE7257CB3196C600 CRC64;	
	Query Match	2.98; Score 195; DB 1; Length 1541;	
	Best Local Similarity	19.58; Pred. No. 0.027;	
	Matches	275; Conservative 163; Mismatches 533; Indels 442; Gaps 68;	
QY	16	LYNSAVSGFDYTIKLTNDQY-----QALNGQLQVLLRF--SKAADGNPNDGNTVDQ 65	
DB	222	LNHEVGG--NNLKLVGDYATYGIAGTPYKVHNNGLIGRGNKEHSOPKG--ILSQD 277	
QY	66	FSKNVATGGN---FDYVKVNG-----NQVEFSGWH-----TQNSDKD 102	
DB	278	PLTNTAVLGDGSPLEFVYDREKRWLFSGYDFWAGYNKKSQWEMNIYKSOFTKDVLNKD 337	
QY	103	SOWIIV-----LVNGKEVRKQ-----LYND 122	
DB	338	SAGSLGKTDYKWSNGKSTITIGGEKSLAVDLADGDKPNHGSVTFEGSGILTNN 397	
QY	123	TKEGAAGFNRRNDYKVNPAIENSMGFGIITL---PVTYKNNVQLVHRFSNDVKTGE 179	
DB	398	IOGAGGLFFEGDYEVKGTSDNTWKG--AGVSARGKTVTKVHNPO---YDLAKLIGK 452	
QY	180	GNVY-----DFWSELMFVDFQKNGP--LKQFGLQITNGQYIIDTTPGPKNFLLQ 232	
DB	453	GTLIVEGTGD-----NKGSLKVGDTGVLKQ---QT--NGSGQHAFASVG-----IV 494	
QY	233	SGNNWYIPDSDFGVGTNALELOFAGTGYSSNBOYRNGAAYSDOKSIENVNGYILTADTW 292	
DB	495	SGRSTLVLDKQVDPNSIYFGRGRID-----LNGN-SLTFD--HRIID-----538	
QY	293	YRPQKILKDGTTWDSKETDMPIL-----HYWNPNTLT-----QAYLYNMKQH 337	
DB	539	-----DGAHLVNNHNTNASNITITGESLITDPTITPYNIDAPDEONPYAFARIKDG 590	
QY	338	GMLPSALPFFRADAPALNHS-----EIVQNIETKRISERTGNTDL-----RT 383	
DB	591	GOL-----YLMLE---NYTYALKGASTRSELPKNGSES--NNWLYMGKTSDEAKN 639	
QY	384	LMBDFVTNPM-----WNKDSNVAFSGTQFGGSL-----KYENSDLTPYANS 427	
DB	640	VM-NHNNRMGFGYFEEBCKNNGNLNVTFKGSEQRNELLTGTGTLNGDLTVEKGT 698	
QY	428	DYKLGMPINIKD-----QTVNGQBFLLANDIDSNPVYQAEQLN-----468	
DB	699	LF--LSGRPTPHARDIAGISSTRKDPHF-----AENHEVYVEDDINRNFKATMNTVNGNA 752	
QY	469	WLYLLNFGTITANNDQNFDSVRVDAPDNIDADLMIADQYFNAAYGMSDQAVSNK---525	
DB	753	SLYSGRNVAMITSNITASKNAQVHIGYK---TGDTCVRSDY--TGYVTCCTDKLSKALN 808	
QY	526	-----HINILEDWNN--ADPEYFNKI---GNPQLTMDDDTIKNSLNH--GLSDATNR 569	
DB	809	SENPTNLGRNVNLTESANFVLGKANLFGTIGRSGNSQVRLE---NSHMLTGNSDVHQL 865	
QY	570	WGLDAIVHQSADRENSTE-----NVVTPNYSFVRAHD--NNSDDQI---ONA-----613	
DB	866	DLANGHTHLNSADSNNTVNTLTVNSLSCNGSFYLLTDLNSKQGDVNVVTKSATGNFT 925	
QY	614	--IRDVTGKDYHT--FTFEDQKG-----IDAY---I 638	
DB	926	LQVADKTGEPNNHNETLFDASKAQDRHLNVLGNTVDLGANKYKLRNNGRYDLYNPEV 985	
QY	639	QIXQNSTVKKYNL---YNPASAYAILLNKDTIPRVYVGDLYTDGGQYMEHQTRYDYDLTN 695	
DB	986	EKRQVTDVTTNITPPNNIQADVPSPNNEELARVDEAPVPPAPAPSETT---ETVAE 1042	
QY	696	LLKSRVYVAGGQSMQTVSGGNNILTSVRYGKGATAT-----DTGTDFT--742	
DB	1043	NSKQESKTVEKNEQDAYETTAQNRREVAKEAKSNYKANTQTMENVAQSGSEKETTETKE 1102	
QY	743	-----RTQGIQGVVYSNTPNKLGLVNDKVVLMHGAHKNQYRAAVLATTD 787	

DB	1103	TATVEKEEKAKVETEKTEQVPKVTSQV--SPKQBSQSETVQPOAEPARENDP--TVNIKEPQ 1159	
QY	788	GVINYNSDQGPVAMIDENGDLYLSSHNLVWNGK---EADTAVQGVAMPDYSYLA---841	
DB	1160	SOITNTTADTQPAKETSSNRYEQPVTESTVNTGNSVVENPENTTPATTOPTVNSSESNKP 1219	
QY	842	-----VWVPVGASDNDQARTA-----PSTEKNSGNSAYRTNAAFDGNVIFEAF 884	
DB	1220	KNRHRSVRSPVHNVEPATSSNDRSIVALCDLTSTWNAVLSDARAKAQFVALNVCKAV 1279	
QY	885	SNEYPTTKESK--RANVYRIAGNADFFASLGFTSFEMAPQYNSSKDRTFLDSTIDNGVAF 943	
DB	1280	SOHISQLEMMNEGYNVWVSNTS-----MMKNYSSSQYRFSK-----1318	
QY	944	DRYDLCHSEPKNYCTDEDLNAIQALHKAGLQVMAQVPPQIYLNLPCKEVATVTRVDDRG 1003	
DB	1319	-----STOTQLGMDQTTISNNVO-----LGQVFTYVRNSNFDKATSKNTLAQV 1361	
QY	1004	NWKDAIINNLYVYNTIGGGEYQKYGGAFLDKQLKLYPEIFTKKQVSTGVAIDFSQKI 1063	
DB	1362	MFYSKYADNHWYLGIDLYGKFKQ-----SKLQTNHNAKFAHHTAQFGLT-----1406	
QY	1064	TEWSAKYFNGPN-----ILHRGSGYVLKADG-----GQYVNL 1095	
DB	1407	---AGKAFNLGNFGITPIGVYRYSYLSNADPQALDQARIKVNPISVKTAFAQVQLSYTHL 1463	
QY	1096	GTYTKQFLPIQLTGEKKQGNISFVKNGDNYTF 1128	
DB	1464	G-----EFSVTPILSARYDANOGSGKINVGIDF 1492	
RESULT 23			
PIP_LACLC	STANDARD;	PRT; 1902 AA.	
AC	PIP_LACLC	PI6271;	
DT	01-APR-1990 (Rel. 14, Created)		
DT	01-APR-1990 (Rel. 14, Last sequence update)		
DT	01-NOV-1995 (Rel. 32, Last annotation update)		
DE	PI-type proteinase precursor (EC 3.4.21.-) (Wall-associated serine proteinase).		
GN	PRTP.		
OG	Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).		
OG	Plasmid pWV05.		
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;		
OC	Lactococcus.		
OX	NCBI_TaxID=1359;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-WG2;		
RX	MEDLINE=88149035; PubMed=3278687;		
RA	Kok J., Leenhouts K.J., Haandrikman A.J., Ledeboer A.M., Venema G.;		
RT	*Nucleotide sequence of the cell wall proteinase gene of		
RL	Streptococcus cremoris Wg2";		
CC	Appl. Environ. Microbiol. 54:231-238(1988).		
CC	-!- FUNCTION: PROTEASE WHICH BREAKS DOWN MILK PROTEINS DURING THE GROWTH OF THE BACTERIA ON MILK.		
CC	SPECIFICITY, ALTHOUGH SOME SUBSITE PREFERENCE HAVE BEEN NOTED, E.G. LARGE HYDROPHOBIC RESIDUES IN THE P1 AND P4 POSITIONS, AND PRO IN THE P2 POSITION. BEST KNOWN FOR ITS ACTION ON CASEINS, ALTHOUGH IT HAS BEEN SHOWN TO HYDROLYSE HEMOGLOBIN AND OXIDIZED INSULIN B-CHAIN.		
CC	-!- SUBCELLULAR LOCATION: Type I membrane protein. Cell wall.		
CC	-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE SUBTILASE FAMILY.		
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).		

```
CC EMEL; M24767; AAA17677.1; --
DR HSP; P00782; L501.
DR MEROPS; S08.019; --
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR003137; PA.
DR InterPro; IPR002029; Peptidase_s8.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR Pfam; PF02225; PA; 1.
DR Pfam; PF0082; Peptidase_s8; 3.
DR PRINTS; PRO0723; SUBTILISIN.
DR PROSITE; PS00136; SUBTILASE_ASP; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; 1.
KW Hydrolase; Serine protease; Cell wall; Zymogen; Signal; Plasmid;
KW Transmembrane.
FT SIGNAL 1 33
FT PROPEP 34 187
FT CHAIN 188 1902
FT DOMAIN 188 1876
FT TRANSMEM 1877 1895
FT DOMAIN 1896 1902
FT ACT_SITE 217 217
FT ACT_SITE 281 281
FT ACT_SITE 620 620
FT DOMAIN 1867 1872
FT SEQUENCE 1902 AA; 199910 MW; 2901C7F1942E5D0B CRC64;
Query Match 2.9%; Score 195; DB 1; Length 1902;
Best Local Similarity 18.7%; Pred. No. 0.036;
Matches 277; Conservative 183; Mismatches 512; Indels 512; Gaps 75;
QY 12 VPSLNSAVSFDITIKNTQYQALNQLQVILLRFSKAAQGNPSGDMTVTD----- 64
DB 316 VFTNSDTSATSTSLVSAIEDSAKIGADVLNMSL-----GSDSGNOTLDEPELAAVQ 368
QY 65 -----QFSKNATGNGFYVKNVNGQVFSFGWHAFTN-----QSNQ 100
DB 369 NANESETAAVISAGNSGTSAGTEGYNKYDYLQDN--EMVGTPTGSRGATTVASAEND 426
QY 101 KDSQWILVNGKEVR-----QLVNDYVEGAAGFNRNDYKVPAINENSMGGF----- 150
DB 427 VITQ-AVYITDGTGLQPGCTIQLSSNDFGS--FDQKKFYVVKDASGNLSKCALADYTA 483
QY 151 --QGIITLVKNEVQLVHRSNDVKTGEGNVDFWSELMPVKDSFQKNG----- 201
DB 484 DAKGKTAI-----VKGELSFDD-----KQYKQAAGAGLIIV 517
QY 202 -----PLKQFGLQTNGQQYYIDPTTGQPRKNFLQSGNNWIYFSDTG 245
DB 518 NNDGTATPTVSMALTTFTPTGLSSVTGKLV-----DWTVAHPDDS 559
QY 246 VGTN-ALELAFAKTGSNQYRNGNAAYSVDKSIDNVNGYLTDATWYRPKQLKGGTT 304
DB 560 LGVKIAL-----TLVFNOKYT-----EDKMSDFTSYGPVSNLSEKPDITAPGGNI 604
QY 305 W-----TDSKETDREPILMWNPNTLQAYLYNWKQHGILLPSALPFFNADADPAE 356
DB 605 WSTQNNNGYTNMSTGSMASPEIAGSALLKQA-----LNKNN-----PFY----- 645
QY 357 LNHYSIVQQNIEKRISGTNTDMLRTMLHDFVTNNPMMKDSNVNFSGI-----QFQGG 412
DB 646 -AYY-----KQLGATLTDLTKIV--EMNTAQPI-----NDINNVNIVSPRQGA 688
QY 413 FL-----KYNSDLTTPYANSYRLLGRPMINKQQTGQGFLLANDIDNSNPTVQA 464
DB 699 GLVDYKAAIDALEKNPSTVAENGY-----PAVELKQDFTSTDKTKLT--FTNST----- 736
QY 465 EQLWLYLLHFGT-----ITANDP-----QANFDSVRVDAPDNTDADLMLNIAQDVEN 512
```

```
DB 737 --THELYQMDSTNDTNAVYTSATDPNSGVLYDKKIDGAIRKAGSNITVPAGKTAQIEFT 794
QY 513 AAY-----GMDSDAVSNKHINILEDNHAD--PEYFNKI-----GN-- 546
DB 795 LSLPKSFDQOQFVEGFLNFKGSDGSRNLNLPYMGFFGGDNDKIVDSLNGITYSPAGNFG 854
QY 547 --PQLTMDDTIKNSLNHGLSDATNRKGLDAIVHQSIA--DRENNSTENVYIPNYSFVRAHD 603
DB 855 TVPLLTNNKNTGTQYQGMVTDAD---GNQTVDDQALAFSSDKNALYNDISKYLLRNIS 911
QY 604 NNSDQIQ---NAIRDVYTKGDYHTFTPEDEQKGDAYIQDNSTVTKKYNLYNPASYAIL 660
DB 912 NVQVILDGQGNKYTLLSSSTNLTKTY-----YNAHSQQYIYNAPAMDGT 958
QY 661 LTNKDFIPRVYIGDLYT-DGQYNEHQTRYDILTNLLKSRVYV-AGCSMQTMSVCGN 718
DB 959 YDQRD-----GNIKTADGYS-----TYRISGVPEGDKRQVDFVFFK 996
QY 719 -NNILTSVRYGKGAMTATDGTDETRIQ-----GIGVVVSTPNLKLGVNDKV 765
DB 997 LDKAPTVRH-----VALSAKTENCKTQYIYTAEKDDLGLDA---TKSVKTAINE-- 1045
QY 766 VLMGAAHKNQOYRAAVIATTTDGVINYT-----SDQAPVAMTDNGDILYSSHNLVV 818
DB 1046 VTNLDAFTDAG-----TTADG---YTKIETPLSDQQAQALGNGDSAEYLYTDN---A 1093
QY 819 NGKEADTAVQGYANPDVSGYLAIVVPVGSNDQDARTAPSTERNSGNSAYR----- 870
DB 1094 SNATDODASVQ---KPGSTSFGLVNGGGIPDKISSTTTGTEANTQGGYTFSGTTPAA 1150
QY 871 -----TNA-----AFQSNVIFEAFSNFYVPTTKESEANVRIAQNAADFFASLGFTS---FE 918
DB 1151 VDCYVYNAQKKHDLNTTYDRAIN---SFTASMPVTNADYAQVDLYAKAHTQLLKHF 1207
QY 919 -----MAP-----QVNSKDRFLDSTI-----DNGYAFTRDYDLGMS 952
DB 1208 TKVRLTAPTETDLKFNNGSDQT-SEATIKVTGTVSADTKTVNVGDTVAALDAQHHFSVDV 1266
QY 953 PNKYG-----TDEDLRNAIQALHAKGLQVNAQVDPQIYNL----- 988
DB 1267 PVNYGNTIKVATDEDDGNTTTEQ-----KITSSYDPMKNSVTFDQGVTFGANETNA 1321
QY 989 -----PQKEVATV-----RVDDRGVWVKDAIINNLYVNVNTIGGEGYKRYG 1031
DB 1322 TSAKFVDPKGTATITGRVKVHTPTILOVDGKQIPKIDDLTFSTLDLGLG---QKPF 1377
QY 1032 GAFLDKLQ-KLYPEIFTKQVSTGVAIDPSQKITWSAKYFNGTNIHRGSGYVLKADGG 1090
DB 1378 VVVGDTTQNKTFQEALT--FILDAVAPTLSLESSTDAPVYITNDPNFQITGTA---TDNA 1431
QY 1091 QYVNL---GTTTKQFLPIQLACEKKQKQNEGFVKGNQVYFVDLAGNMVKNFTIEDSVG 1146
DB 1432 QYLSLSINGSSVASQVVDINISGK-----PGHMA----- 1461
QY 1147 NWYFFDQDGKAMENKHFVDV---DSYGEKGTFFYFLKNGSVFRGLVQTDNGTYFDNYGK 1203
DB 1462 ----IDQPKLLSEGNVLVAVTDS--EDNT-----TTKNITVYE---- 1496
QY 1204 MVRNQTINAGAMIYTDENGLKLIKASYNDAEYPT---STDVCK 1244
DB 1497 --PKTLAAPTVPSTTEPAKTVTLTANSAAATGETVQYSADGK 1538
RESULT 24
YMYB_CABEL STANDARD; PRT; 918 AA.
ID YMYB_CABEL
AC P34487;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Hypothetical 96.6 kDa protein F59B2.12 in chromosome III.
GN F59B2.12.
OS Caenorhabditis elegans.
```

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
OC NCBI_TaxID=6239;
RN SEQUENCE FROM N.A.
RP STRAIN=BRISTOL N2;
RC MEDLINE=94150718; PubMed=7906398;
RX Wilson R., Ainscough R., Anderson K., Baynes C., Berks M., Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A., Craxton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L., Jones M., Kershaw J., Kirsten J., Laister N., Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M., Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkeen R., Sims M., Smalton N., Smith A., Smith M., Sonhammer E., Staden R., Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans."
RL Nature 368:32-38(1994).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC
DR EMBL; Z11505; CAA77381.1; -;
DR PIR; S31132; S31132.
DR Wormpep; F59B2.12; CE01024.
KW Hypothetical protein.
SQ SEQUENCE 918 AA; 96560 MW; E464FD86B14945DE CRC64;

Query Match 2.9%; Score 191; DB 1; Length 918;
Best Local Similarity 18.2%; Pred. No. 0.021;
Matches 210; Conservative 173; Mismatches 456; Indels 318; Gaps 51;

QY 47 RPSKADGNPSGNDTVDQSKYATGNGEDYVKNQGV-----EFSGWHDATQSN 99
DB 29 KFSSSSG--FGASMFSSNGAK--SSFGGFAAMPKLDASKAAVHSSSHKHGQSSGS 84
QY 100 DKDSQMLIVLVNGKEVKRLQVNDKEGAGFNRNDYKVPATENSMSGFGIIFLPTV 159
DB 85 SYNTHSLTVVGADSKN-----ITENSEKKDYKNES--KVDEANENTKIRKSDGSV----- 132
QY 160 VKNENVOLVHRFSNDVKTGNGVDFWSELMVPVKDFQKNGPLKQFGLQTNGQOYYID 219
DB 133 -----IETGK-----SHNKSDDASSYGLE--KSKRTYAD 160
QY 220 PTTGQPKNFLLQSGNWIYFDDTDGVTGNALELQFAKGVSSNEQYRNGNAYSYDDKS 279
DB 161 -----KNTMTLSSN-----TKINKQSRSAALDEGNEFVQNA--DGTG 198
QY 280 IENVGYLTADTWYRPRKQILKDGTTTDSKETDRPILAVVWNPNTLTQAYLYNMQHGN 339
DB 199 LRNTGK-----KNTD-----EHLSH----- 214
QY 340 LLPALPFFNADAPAEINHYSEIVQONIKRISYEWNTDRLTLMH--DFVTNNPMNKD 398
DB 215 -----NYLDENAOQMSIGADGTSNITNRKSGVSDSHNAASDAHSNFSLSDAQNGK 265
QY 399 SENYFSCIQGQGLAYENSDELTPYANSYRLGLRMPINKDTYRGQEFELLANDINS 458
DB 266 SONYSKKAASAGSNADE--SNLESKNADGTSMSNGFNNTSY-----DKATA 315
QY 459 NPVVQAEQLNWLTYLLFGTITANNQANFDSVRDAPDNIDADLMIAQDYFNAAYGMD 518
DB 316 EVVSKKNVN-----ADGTSMEASHAGSNSSKINSASGSGSLSLVGPN-----GIK 363

QY 519 SDAYSNNKHIN--ILEDNNHADPEYFNKIG-NPOLTMDOT-----IKNSLNHGLSDAT 567
DB 364 SHSTSNKTDNIALDEANQSGSISEQIGKNGQSLNESSIESQKRAESRNTADLDSV 423
QY 568 NRMGLDAIVHQSLSA-----DRENNSTENVVPIPNYSFVRAHNNNSQOIQAIHARDVTGKDY 622
DB 424 DANGTVSSSHSKSASGTSLDENHNTKHAL-----QASVDEHGNMKNHSDIGSYRNKK----- 475
QY 623 HTTFEDEQGIADYIQDQNSQTVKKYNLYNIPASYALLLNKDTIPRVYVYDLYTDGGQY 682
DB 476 ---TGEFGNSSENSASIKNADGTMQSVNVKND-----TNRMT-----YEAKSA 515
QY 683 ME--HQRYYVDLTNLIKSRVYVAGQSMQTMVSGGNMILTSVRYGKGAMTATDITGD 740
DB 516 LEKHNKNSDGTFKDESK-----GNSRVRNTDGGSNLAVGSYVKGKGVSSNET--- 565
QY 741 ETRTQGGVYVVSNTFNLKLVNKKVLMGAAHKNQYRAAVLTTTDDGVINYSDDQAPV 800
DB 566 -----IASSNAFN-----TSDAESNQFDHLHOKTANGTEITHAKDSKQVA 605
QY 801 AMTDENGDLILSSHNLVYNGKEADTAVQGVANPDVSGYLAVVVPVCGASDNOQART---A 857
DB 606 ASANAKSLDTSMSAVDAKGNKVDKTSQAADSHDA-----ISASDQVDAKIYKHA 656
QY 858 PSTKNSGNSAYRTNAAFDNSVIFEAFSNFYITPKESRANVRFAQNADPFASLGTFSF 917
DB 657 DRSESISNDNSQVASEHNS-----SKQSEHEK---RQNA--GSFSDVSS 698
QY 918 EMAPQYNSKDKORTFELDTIDNGYAFTRDYDLGMBEPKNKYGTDELRNAIQALHAKAGLVPM 977
DB 699 NSAKK--NAVKDAT---DVRQNAFANVDAAAGNSVSE-----VKNSLVENHNSDAS 745
QY 978 ADWVPDQIY--NLPGKEVATVTRVDRDGNVKNKDAIINNLYVNTIGGGEYQKKY---G 1031
DB 746 SD---SKTHSKSVDTGTELTDA-----KHSNVSTSHLTEGQL-----AQOKKLAMADG 790
QY 1032 GAFDLQKQLYPFIPTKQVSTGVVAIDFSQ---KITWSAKYFNCTNITLHRSGSYVLKAD 1088
DB 791 SVVMSNDNSHITHEKSRADVADAGHAHLKADGSGSTIDEGFKHSDLESRGEG---AQ 846
QY 1089 GQGYNLGTTTQKFIPIQLTGEKKQGNESFVKGNDGNYFYFDLAGNMYNKTFIEDSVGNW 1148
DB 847 KORYQKLGNGTSSMDV---GYEK---SMAKGGDOT-----SSHKKTLAKDGKGH- 890
QY 1149 YPFDQDGKMKVENKHFVD 1165
DB 891 FTETKDG--ESHKKID 905

RESULT 25
ICAO_HAEIN
ID ICAO_HAEIN STANDARD; PRT; 1694 AA.
AC P44969;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Immunoglobulin A1 protease precursor (EC 3.4.21.72) (ICAI protease).
GN ICA OR ICAI OR H10990.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SEROTYPE D;
RA Wright A., Fisman Y., Tai F., Plaut A.G.;
RL Submitted (MAY-1991) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F., Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

DM protein - protein search, using sw model

Run on: August 12, 2002, 07:53:59 ; Search time 46.36 Seconds
(without alignments)
4668.176 Million cell updates/sec

US-09-995-749a-2_COPY_531_1781

Title:

Perfect score:

Sequence:

Scoring table:

Gapop 10.0 , Gapext 0.5

Searched:

562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters:

562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

SPTREMBL.19.*

1: sp_archaea.*

2: sp_bacteria.*

3: sp_fungi.*

4: sp_human.*

5: sp_invertebrate.*

6: sp_mammal.*

7: sp_mhc.*

8: sp_organelle.*

9: sp_phase.*

10: sp_plant.*

11: sp_protein.*

12: sp_virus.*

13: sp_vtebrate.*

14: sp_unclassified.*

15: sp_rvirus.*

16: sp_bacteriap.*

17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2690	40.5	2057	2 Q9RE05	Q9re05 leuonostoc
2	2546.5	38.3	1527	2 Q9ZAR4	Q9zar4 leuonostoc
3	2474.5	37.3	1477	2 Q9L466	Q9l466 leuonostoc
4	2467	37.1	1508	2 Q52224	Q52224 leuonostoc
5	2465	37.1	1508	2 Q9E2H5	Q9ezh5 leuonostoc
6	2375.5	35.8	1577	2 Q54178	Q54178 streptococ
7	2371	35.7	1449	2 Q68542	Q68542 streptococ
8	2371	35.7	1449	2 Q55264	Q55264 streptococ
9	2346.5	35.3	1455	2 Q69397	Q69397 streptococ
10	2343.5	35.3	1455	2 Q69388	Q69388 streptococ
11	2342.5	35.3	1455	2 Q69382	Q69382 streptococ
12	2341.5	35.3	1390	2 Q69385	Q69385 streptococ
13	2340.5	35.2	1455	2 Q69391	Q69391 streptococ
14	2330	35.1	1575	2 Q9LCH3	Q9lch3 streptococ
15	2289	34.5	1590	2 Q55263	Q55263 streptococ
16	2284	34.4	1590	2 Q59983	Q59983 streptococ

17	2256	34.0	1290	2	Q48756	Q48756 leuonostoc
18	2180	32.8	1338	2	Q9WX14	Q9wx14 streptococ
19	2175.5	32.8	1577	2	Q55265	Q55265 streptococ
20	2119.5	31.9	1512	2	Q9WXJ5	Q9wxj5 streptococ
21	2093	31.5	1518	2	Q00600	Q00600 streptococ
22	2005.5	30.2	1599	2	Q00599	Q00599 streptococ
23	1989	30.0	1016	2	Q9LQJ7	Q9lcj7 leuonostoc
24	317	4.8	2817	16	Q97K42	Q97k42 clostridium
25	251.5	3.8	1385	16	Q99M47	Q99m47 staphylococ
26	246	3.7	2893	16	Q25063	Q25063 helicobacte
27	241.5	3.6	601	2	Q45821	Q45821 clostridium
28	241.5	3.6	1315	2	Q86488	Q86488 staphylococ
29	239	3.5	2902	16	Q9ZME6	Q9zme6 helicobacte
30	229.5	3.5	3194	16	Q9ZLM3	Q9zlm3 helicobacte
31	228	3.4	1395	2	Q9A1S0	Q9a1s0 staphylococ
32	225.5	3.4	2364	2	Q46342	Q46342 clostridium
33	222.5	3.4	1844	2	Q96201	Q96201 plasmodium
34	219	3.3	1604	2	Q9KK99	Q9kk99 rickettsia
35	217.5	3.3	1806	16	Q929J3	Q929j3 listeria in
36	216.5	3.3	1463	2	Q86919	Q86919 staphylococ
37	215	3.2	2340	16	Q9ZD91	Q9zdg1 rickettsia
38	214	3.2	1181	5	Q97260	Q97260 plasmodium
39	214	3.2	6713	16	Q99054	Q99054 staphylococ
40	214	3.2	6713	16	Q931R6	Q931r6 staphylococ
41	213.5	3.2	1377	16	Q98QZ8	Q98qz8 mycoplasma
42	213	3.2	1227	16	Q97K41	Q97k41 clostridium
43	213	3.2	2399	16	Q9ZKS9	Q9zks9 helicobacte
44	211.5	3.2	1849	2	Q9S4K2	Q9s4k2 lactobacill
45	211	3.2	1578	16	Q92E25	Q92e25 listeria in
46	211	3.2	2367	2	Q46034	Q46034 clostridium
47	210	3.2	1939	16	Q97DT3	Q97dt3 clostridium
48	210	3.2	2367	2	Q9F931	Q9f931 clostridium
49	209	3.1	4152	2	Q9ZHL3	Q9zhl3 haemophilus
50	209	3.1	4688	16	Q9PQ08	Q9pq08 ureaplasma
51	209	3.1	4919	2	Q9ZHL0	Q9zhl0 haemophilus
52	208.5	3.1	1615	2	Q9KKA4	Q9kka4 rickettsia
53	208	3.1	2353	2	P71401	P71401 haemophilus
54	207.5	3.1	583	2	Q45822	Q45822 clostridium
55	207.5	3.1	1983	16	Q9CJ55	Q9cj55 lactococcus
56	207.5	3.1	2367	2	Q9EXR0	Q9exr0 clostridium
57	206.5	3.1	5005	16	Q9PP25	Q9pp25 ureaplasma
58	205.5	3.1	1564	2	Q93IM3	Q93im3 staphylococ
59	205	3.1	1361	2	Q87083	Q87083 wolfinella r
60	205	3.1	2295	5	Q9TY98	Q9ty98 plasmodium
61	204.5	3.1	2529	16	Q25579	Q25579 helicobacte
62	203.5	3.1	1295	2	Q34110	Q34110 helicobacte
63	202.5	3.0	1557	2	Q9BN12	Q9bn12 haemophilus
64	202.5	3.0	1819	16	Q97K40	Q97k40 clostridium
65	202.5	3.0	2167	16	Q92EK5	Q92ek5 listeria in
66	202	3.0	1139	5	Q97237	Q97237 plasmodium
67	201.5	3.0	1217	16	Q97GW3	Q97gw3 clostridium
68	201	3.0	1403	2	Q44494	Q44494 azotobacter
69	201	3.0	1883	16	Q9PPT2	Q9ppt2 ureaplasma
70	200.5	3.0	486	2	Q68875	Q68875 streptococ
71	200.5	3.0	1938	2	P70983	P70983 bacillus sp
72	200	3.0	961	2	Q9F2D9	Q9f2d9 salmonella
73	200	3.0	1645	16	Q97MX7	Q97mx7 clostridium
74	200	3.0	2894	5	Q96223	Q96223 plasmodium
75	200	3.0	2500	17	Q58791	Q58791 methanococ
76	199.5	3.0	1305	2	Q9EZE7	Q9eze7 escherichia
77	199.5	3.0	1733	2	Q9K114	Q9k114 staphylococ
78	199	3.0	1288	2	Q93UM8	Q93um8 helicobacte
79	199	3.0	1616	2	Q9KKA2	Q9kka2 rickettsia
80	199	3.0	1616	2	Q9KKA1	Q9kka1 rickettsia
81	198.5	3.0	953	2	Q48502	Q48502 lactobacill
82	198.5	3.0	3890	16	Q99U53	Q99u53 staphylococ
83	198	3.0	949	5	Q97306	Q97306 plasmodium
84	198	3.0	1132	16	Q9PQ35	Q9pq35 ureaplasma
85	198	3.0	1306	2	P77070	P77070 escherichia
86	198	3.0	4199	16	P74440	P74440 synechocyst
87	197.5	3.0	1616	2	Q9KKB8	Q9kbb8 rickettsia
88	197.5	3.0	1643	2	Q9F0P7	Q9f0p7 rickettsia
89	197	3.0	635	2	Q45820	Q45820 clostridium

90 197 3.0 694 16 Q97BM5 Q97em5 clostridium
91 196 3.0 1144 2 Q50371 Q50371 mycoplasma
92 196 3.0 1616 2 Q9KKA9 Q9KKA9 rickettsia
93 195.5 2.9 1643 2 Q9F0P6 Q9F0P6 rickettsia
94 195 2.9 1238 16 Q25330 Q25330 helicobacte
95 195 2.9 1603 2 Q9KKA6 Q9KKA6 rickettsia
96 194.5 2.9 1092 2 Q70022 Q70022 staphylococ
97 194 2.9 1260 16 Q931M2 Q931M2 staphylococ
98 194 2.9 1616 2 Q9KKA7 Q9KKA7 rickettsia
99 194 2.9 1758 16 Q9JMS5 Q9JMS5 escherichia
100 193 2.9 533 2 Q9AQ54 Q9AQ54 bacillus me

ALIGNMENTS

RESULT 1
Q9RE05 PRELIMINARY; PRT; 2057 AA.
ID AC Q9RE05;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE ALTERNANSUCRASE (EC 2.4.1.140).
DE ASR.
GN Leuconostoc mesenteroides.
OS Bacteria: Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
OC Leuconostoc.
OX NCBI_TaxID=1245;
RN [1]
RS SEQUENCE FROM N.A.
RC STRAIN=NRRL B-1355;
RX MEDLINE=20080809; PubMed=10612736;
RA Arguello-Morales M.A., Rемаud-Simeon M., Pizzut S., Sarcabal P.,
RA Willemot R.M., Monsan P.
RT "Sequence analysis of the gene encoding alternansucrase, a sucrose
glucosyltransferase from Leuconostoc mesenteroides NRRL B-1355."
RL FEMS Microbiol. Lett. 182:81-85(2000).
RM EMBL; AJ250173; CAB65910.2; -
DR InterPro: IPR006604; Chlamydia.OMP.
DR InterPro: IPR002479; CW-binding.
DR InterPro: IPR003318; Glyco_hydro_70.
DR Pfam: PF01473; CW-binding_1; 12.
DR Pfam: PF02324; Glyco_hydro_70; 1.
DR ProDom: PD001717; Chlamydia.OMP; 1.
DR Transferase: Glycosyltransferase.
SQ SEQUENCE 2057 AA; 228987 MW; 62BC59385D9A11BE CRC64;

Query Match 40.5%; Score 2690; DB 2: Length 2057;
Best Local Similarity 45.7%; Pred. No. 2.1e-115;
Matches 604; Conservative 175; Mismatches 385; Indels 158; Gaps 40;
QY 30 LTNDQYALQALQVLLRESKAADGNPSGNTVTDQFSKNVATGGNFYKVNQGVF 89
DB 153 ITGSHYNNHGYFVVIDASGQYVGLQNIDGNL-QYFDDNCGYQVKGSR-DYNGKHIYF 209
QY 90 SGWHTAQSDNDKQWIIIVLNGK-----EVKRLVNDTKEGAAGTNRNDYKVP 140
DB 210 DSVTGKASN-----VDIVNGKAQGYDAQGNLKKSYVADSSGQYFYFDGNG----QP 258
QY 141 ALENSM-----SFGQIILPVTVKKNVGVLRVPSNDYKTGE-----179
DB 259 LIGLOTIDGNLQYFNQGVQVIGGFQ-----DVNKKRIYFAPNTGNVANTETIINGKL 311
QY 180 -----GNVDFPFWELMPKDSFK-----GNCPLKQFGLQTINGQVYIDPTTG 223
DB 312 QGRDANGN-----QVKNRFSKDVAGNTFYFDANGVMT-GLQISGKTYILD-EQG 360
QY 224 QPKNFILQSGNNWTFDSGCVGTNALEQAFKGTYSNNGQYRNNGNAVSYDDKSIENY 283
DB 361 MLRKNVAGTFNNQFMVFDADTGAGKTALEYQFDQGLVGSQSNENTPHNAKSYDKSSFENY 420

QY 284 NGYLTAADWYRBPQKILKDGTTWDSKSTDMRPIILMVWPNLTLQAYLYLNYKQGNLLPS 343
DB 421 DGYLTADWYRPTDILKNGDWTATSTETDMRPLLTMTWTPDQQTQANTLNFNSKGLGIT 480
QY 344 ALPFFNADADPAELNHYSEIYQONIEKRISTGTMDLKRILMHDFVTNNPMWKNDSNVN 403
DB 481 T---YTAATSOKTLNDAAFVQTALQQISLKKSTEHLDAIDSFYKVTQANWKNQTEDEA 537
QY 404 FSGIQF-QGGFLKY-ENSDLPYANS-DYRLLGRMPINI---KDOT-YRQGEFFLLANDID 456
DB 538 FDGLQWLQGGFLAYQDDSHRTPNTOSGNRKLGRQPINIDGSKDTTDGKSEFFLLANDID 597
QY 457 NSNPVQAEQNLWLYLLNFGTITANNDQANFDSYRVDPADPNIDADLNIADQYFNAAYG 516
DB 598 NSNPVQAEQNLWLYLLNFGSITONNDNANPDGRYDAVDVNDADLKLKIAGDFFKALYG 657
QY 517 MD-SDAVSNKHINILEDWNIHADPEYFNKIGNPQLTMDDTIK-----NSLNHGLSDATNRW- 570
DB 658 TDKSDANANKHLSILEDWNGKDPQVYVNOGNAQLTMDYTVTSQFCNSLTHGANNRPNMY 717
QY 571 -----GLDAIVHQSLADREN-----NSTENVVIPNYSFVRAHDNNSDQIQNA 613
DB 718 FLDTGYLLNGDLNKKIVDKNRPNSGTLVNRANSGLTKVIPNYSFVRAHDYDAQDPIRKA 777
QY 614 IED--VTGKDYHTFTEDEQKIDAYIQDN--SVKKYLNLYNIPASVAILLTNNKDTIPR 669
DB 778 MIDHGILIKNMQDTFTFDLAQGEFYKDOENPSGFKYINDYLNLSAYAMLLTNKDTVPR 837
QY 670 VYTGDLTYDGGYMEHQTRYDTLTNLKSRVYVAGGQSGMQTNSVG-----GNNNLTLS 724
DB 838 VYTGDMYLSGGQYMEKGTIYNPVISALLKARIKYVSGQTMATDSSGRDLKDGEDTLTSS 897
QY 725 VYTGKAMTATDTGDEET----RTQIGVYVNTNPKLVNDKVVVLHMGAAHKHQYRA 780
DB 898 VRFGKIMTSDQTTQDSQDYKNOGIGYVGNPNLKLNNDKTITLHMGAKHKQYRA 957
QY 781 AVLTITDGYVINTSDQAPVAMTDENGOLYLSSHNLVYNGKEA---DTAVQGYANPDVS 837
DB 958 LVLSNDSGIDVDSDDKAPLTNDNGDLIFKHTNFFV--KODGIYINWEMKGSUNLALIS 1015
QY 838 GYLAWVVPVGCASDQDARTAPSTKNSGN--SAYRTNAAFDSDNVIFAEFSNFVYPTPKES 895
DB 1016 GYLGWVVPVGCASDSQDARTV-AVESSSSNDGVSFVHSAALDSNVIEGFSNFQAMPTSPE 1074
QY 896 ERANVIAQNADEFFASLGFTSPEMAPQYNSKD-----RTFLDSTIDNGYAFTRDYLGM 950
DB 1075 QSTNVVIATKANLFKELGITSFELAPQYRSSGDTNYGMSFLDSFLNNGYAFTRDYLGF 1134
QY 951 SE-----PNKYGTDEDLRNAIQALHAKAGIQVMADWVPQIYNLPGKEVATVTRVDGRN 1004
DB 1135 NKADGNPNPKYGTDDLRNAIELHNGMQALADWVPQIYALPGKEVTVATRVDERGN 1194
QY 1005 VWKDAIINNLVYVNTIGG-ELYKKYGGAFDLKLOKLEIPIFTKKQVSTGVAIDPSQKI 1063
DB 1195 QLKDDFVNLVYVNTKSSGVYQAKYGGFLLKLEEFVPSLFKQNVSGQPIDASTKI 1254
QY 1064 TEKSAXFNGTNIHRCSGYVLA-KDGGYVNLGTTTKQFLPQLTQGEKKQGNQGVKGN 1122
DB 1255 KQSAKYNGTNIHRCGYVYVLDKWDATNQYFNIAKTNEVEFLPLQL--QNKDAQTGFLSDA 1312
QY 1123 DGNVYFYDLAGNVKNTFIEDSVGNWYFFDQDGKMYENK-----HFVDVDSYGEKGTFF 1176
DB 1313 SGVYK-YISIGYQAKDTFIEDGNGWYFFDKDGYVYVRSQGNPIFTVETSVNTRNGY 1371
QY 1177 FLKNGVSFRGLYQTDN--GTYFVNDYGNKVRNQTINAGA-MITYTLDENGLIKKASYNSD 1233
DB 1372 FMPNGVELRKG-ETDNGSNVYFFDQDGKMYRDKYINDDANNFYHLNVGDTMSRGLFKED 1430
QY 1234 AE 1235
DB 1431 SD 1432

593 IPNYSFVRAHDNNSOQONATROV--TGKDYHTTFFDEDEKGDIDAYIQDNSTVVKYNL 650
Db 634 IPNYSFVRAHDSEVQTVIAIISLHPDVKNLSLAPLADQLABAFKIYNNDEKQADKTYTQ 693
QY 651 YNPASAYALLTNKOTIPRVYVGYDLYTDGGQYMEHQTRYIDFTLNLKSRVKYVAGGQSM 710
Db 694 YNMPASYALLTNKOTIPRVYVGYDLYTDGGQYMEHQTRYIDFTLNLKSRVKYVAGGQSM 753
QY 711 QTMVSGNNILTSVRYGKGMATADTGTDETQTCIGVVSNTPLNLKLGVDNKKVYLHMG 770
Db 754 AV---DQNDILTNRYGKGMATADTGTDETQTCIGVVSNTPLNLKLGVDNKKVYLHMG 809
QY 771 AAHKNOQYRAAVLTVTDGQVINTSDGAPVAMTDEGDLVLSHNLVYNGKEADTAVQG 830
Db 810 AAHKNOQYRAAVLTVTDGQVINTSDGAPVAMTDEGDLVLSHNLVYNGKEADTAVQG 856
QY 831 YANPDVSGVYLAWVVPVGA SDNDQARTAPSTEKNSGNSAYRTNAFDSNVIPAFSNFYVT 890
Db 857 VRNPQVSGVYLAWVVPVGA SDNDQARTAPSTEKNSGNSAYRTNAFDSNVIPAFSNFYVT 916
QY 891 PTKESRANVRIAQNAADFPASLGTTSFEMAPQYNSKDRFLDSTIDNGYAFTRDYDLM 950
Db 917 ATNTEDYTNAVIAKNGQLPKDNGJUSFQLAPQYRSSTDPFLDSIIQNGYAFTRDYDLY 976
QY 951 SEPKNYGTDEDLNNAIQALHKLQVMAADWYFDQIYNLPKGEVATVTRVDRGNVWKDAI 1010
Db 977 GTPKTYGTVDQLDAIKALHKLQVMAADWYFDQIYNLPKGEVATVTRVDRGNVWKDAI 1036
QY 1011 INNNLVVNTIGGGEYQKYGGAFLDKLQKLYPEITKKOVSTGVAIDPSQKITMSAKY 1070
Db 1037 IDGSLVQSGGKQYQAGYGAFLDIQKYPALFETQIISTGLPMDPSQKITMSGKY 1096
QY 1071 FNGTNIHLRSGYVKADG-GOYNNL--GTTTKQFLPQLTGEKKQGNKPGVKGNDGNY 1127
Db 1097 FNGSNIQKGAGYVKADGSDQYKVTNNNNRDELPLQLTDDLSR--FVRDNIQWY 1154
QY 1128 FYDLAGMKNFTEDSVGNWYFFEDODGKMKVHFDVDSYGEKGYFELKNGYSFRGG 1187
Db 1155 -YTLSGILARNTIQDNGNYYIFDSTGHLVYG--PQNNH---HYFELNGIELWOS 1207
QY 1188 LVQ-TONGTYFDNYGKMYRNOI-----NAGMIYT-----LD-----ENK 1224
Db 1208 FLQVADGSTIYFQKGRVFNQYITDQGTAYTFQNDGTMTVSGFTFIDGHKGYFYKNGT 1267
QY 1225 LKASVNSDAE 1235
Db 1268 QVKGQFVSDT 1278

RESULT 6
Q54178 PRELIMINARY; PRT: 1577 AA.
AC Q54178; Q54247;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DE 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE GLUCOSYLTRANSFERASE.
GN GTFG.
OS Streptococcus gordonii challis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=29390;
RN 111
RP SEQUENCE FROM N.A.
RC STRAIN=CHALLIS;
RX MEDLINE=96157084; PubMed=8586195;
RA Vackerman M.M.; Sulavik M.C.; Clewell D.B.;
RT "Molecular analysis of Streptococcus gordonii glucosyltransferase
phase variants.";
RL Dev. Biol. Stand. 85:309-314(1995).
RN SEQUENCE OF i-96 FROM N.A.
RP STRAIN=CHALLIS;
RC

RX MEDLINE=92276337; PubMed=1534326;
RA Sulavik M.C.; Tardif G.; Clewell D.B.;
RT "Identification of a gene, rgg, which regulates expression of
glucosyltransferase and influences the spp phenotype of Streptococcus
gordonii Challis.";
RL J. Bacteriol. 174:3577-3586(1992).
DR EXBL; U12643; R4C43483.1; -;
DR EXBL; M89776; AAA28969.1; -;
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003118; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 18.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Transferase.
SQ SEQUENCE 1577 AA; 177805 MW; 5AE0328DC5E08D18 CRC64;
Query Match 35.8%; Score 2375.5; DB 2; Length 1577;
Best Local Similarity 47.4%; Pred. No. 4.2e-101;
Matches 493; Conservative 159; Mismatches 328; Indels 59; Gaps 18;
QY 208 LQTINGQYIYDPTTGGQPRKNFLQSGNNWYFDSDTGVGTNALLELOPAKGTVSSNQYR 267
Db 228 IKTIDGKYYVQ-DGQTVKKNFAVELNGKILYFDAETGALVDSNEYQFQGTSSINNEFT 286
QY 268 NGNAAYSYDDKSIENWVGYLTADTWYRPQILKDGTTWDSKETDMRFLVWHPNLTQ 327
Db 287 QKNADYGTGDKDIEVDSGLYTDADSWYRPKFLKDGKTWTASTETDLRLFLMAMWPKRKTQ 346
QY 328 AYLNYMKQHGKLLPSALPFFNADAPAEALNHYISVIOQNTKEKRISETGNTDMLRTLMHD 387
Db 347 INYLVNQE-NL--GIGAFESKTEQVLLTNVQVQVKIEERISKEDTKLRLTMSA 402
QY 388 FVTNPMN--KDSNVNPFSGIQFGGFLKYENSOLTPYANSDYLLGLRMPIN----- 438
Db 403 FYKTOPNNIKTESSETGNTKDLQGGALLYTNDSKTSKANSRYLLNRTPTSGTGPY 462
QY 439 IKDQYRGQEFLLANDINSNPVQAEOLNLYLLNFTTANNDQANFVSVRVADPN 498
Db 463 FIDKSGNGVEFLANDFNSNPVQAEOLNLYLLNFTTANNDQANFVSVRVADPN 522
QY 499 IDADLNTAQDYFNAAYGM-DSDAVSNKHINILEDNNHADPEYFNKIGNPQITMDDTIKN 557
Db 523 VNADELLOIASDYFKSRYKVGSEEEALKHLSLEAWSNDPDYNKDTKGAOLAIDNKLRL 582
QY 558 SLNKG-LSDATNRGLDAIVHQSLADRENNSTENVINYSFVRAHDNNSQDQIONAIRD 616
Db 583 SLLYSFMRKLSIRSGVEPTITSLNDRSTENKERTANYIEFRAHDSVQTVYAOIIRE 642
QY 617 --VTGKDYHTTFFDEDEKGDIDAYIQDNSTVVKYNLXNIPASAYALLTNKOTIPRVYGD 674
Db 643 NINPNTDGLTFTMDLQKAFKIYNEDMRKADKKYQFNIPTAHALMLSNKDSITRVYGD 702
QY 675 LYTDDGQYMEHQTRYIDFTLNLKSRVKYVAGGQSMQMSVG-----GNNNLTSTVR 726
Db 703 LYTDDGQYMEKKSPYHDAIDALLARIKYVAGGQMKVYMGVPREADKNSYNGILTSVR 762
QY 727 YGKGAMTADTGTDETQTCIGVVSNTPLNLKLGVDNKKVYLHMGAAHKNOQYRAAVLTT 786
Db 763 YGKGAMTADTGTDETQTCIGVVSNTPLNLKLGVDNKKVYLHMGAAHKNOQYRAAVLTT 822
QY 787 DGYINTYTDQGAFAVAM---TDENGDIYLSHNLVYNGKEADTAVQGANPDVSGYLAW 843
Db 823 DGYINTYTDQGAFAVAM---TDENGDIYLSHNLVYNGKEADTAVQGANPDVSGYLAW 870
QY 844 VPVGASNDQARTAPSTEKNSGNSAYRTNAFDSNVIPAFSNFYVTPTKESERANVRIA 903
Db 871 VPVGASNDQARTAPSTEKNSGNSAYRTNAFDSNVIPAFSNFYVTPTKESERANVRIA 930
QY 904 QNADFPASIGTSFEMAPQYNSKDRFLDSTIDNGYAFTRDYDLMSEPKYCYDDELRL 963
Db 931 KVNLFKNGVTSFELPPQYVSSQDGFLLDSIIQNGYAFTRDYDLMSEPKYCYDDELRL 990
QY 964 NAIQALHKLQVMAADWYFDQIYNLPKGEVATVTRVDRGNVWKDAIINNLYVYNT-IG 1022

Db 991 NALRALHVSNIQALADVPQIYNULPGKEVVVATRVNNYGYREGAEIKELVYANSKTN 1050
QY 1023 GGEVOKYGGAFDLKLOKLPETFKQVSTGVAIDPSQKITWSAKYFNGTNILRGSG 1082
Db 1051 GTDYQKYGGAFLDELAKAKYPEIFERVOISNGOKMTDEKITKWSAKYFNGTNILRGAY 1110
QY 1083 YVLKADGGQYVNLGTTTKQPLTQLGEGKQKQNEGFVKGNDQNYFYFDLAGNMMVKNFTIE 1142
Db 1111 YVLKDWGSKYLSNKGNETALPKQLV--NKEASTGFVKDPNG-FKYSYSGNAQKDFIQ 1167
QY 1143 DSVGNVYFFDQKQWVKNHFVDVDSVGEKGTFFFLKNGVSFRGGLVQTDNGT-YYPDNY 1201
Db 1168 DENGNYFFEDNQYLVNAGAREID-----GKQ--LYFMKNGVQLRDALQEDENGQYYDYKT 1221
QY 1202 GKMYRNTINAGAMIYTL 1220
Db 1222 GAKVLR-----YVTS 1233

RESULT 7
Q68542 PRELIMINARY; PRT: 1449 AA.
AC Q68542;
DT 01-AUG-1998 (Tremblrel. 07, Created)
DT 01-AUG-1998 (Tremblrel. 07, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE GJCOSYLTRANSFERASE N (FRAGMENT).
GN GFN.
OS Streptococcus salivarius.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1304;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-VI477;
RA Jaffe R.I.;
RT *Streptococcus salivarius VI477 gtfN.*;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF049609; AAC05156.1;
DR InterPro: IPR002479; CW_binding.
DR InterPro: IPR003318; Glyco_hydro_70.
DR Pfam: PF01473; CW_binding_1; 8.
DR Pfam: PF02324; Glyco_hydro_70; 1.
KW Transferase.
FT NON_TER 1449 1449
SQ SEQUENCE 1449 AA: 159895 MW: 0700F6D748471BFB CRC64;

Query Match 35.7%; Score 2371; DB 2; Length 1449;
Best Local Similarity 42.5%; Pred. No. 6.1e-101;
Matches 554; Conservative 170; Mismatches 417; Indels 164; Gaps 40;

QY 5 ARPDVAAYPSLYNSAVSGFD--TTIKLTNDQALNGLOVILLRKSADGNPSGDNTV 62
Db 52 AIPDITVGTGVNDTAAADPTTAAATND---VAPDQATPTAFIDLTDTT-----NTV 104
QY 63 TDQFSKNYATGNG-----FDYKVGNGNOYEEFSGWHATNOSNDKDSQWILVLYNGREVR 117
Db 105 AANAVDTVAIVGTDRRAATNDTATNDTAVD-----TTNNNTTDT---TTVTRRAATTE 156
QY 118 QLVNDTRKEGAAGNRNDVYKVPAINSENSMSGPGGYLTLPVTVKENVOLVHRESNDVKT 177
Db 157 RRATGARRGPTGGR--ATPVNGTNNANT-----VTVNNDLPA-----TNNVVT 201
QY 178 GEGNYVDFWSELMPKDSFKGNGPLKQFLOLQINGQOYVIDTTPGPRANFLLQSGNW 237
Db 202 -----GPPSH-----IKTNGKQYVVE--DXDTIRKNYVLERIGGS 235
QY 238 IYFSDTRGVGTNALEQFAK--GTVSSNEQYR-----NG--NAAYSYDDKS IENVNGILT 288
Db 236 QYFNAETGELSNQKEYRDKNGTGTSSADSTNTNTVTVNGDKNAFYGTGTDKDIELVDGIFT 295

QY 289 ADTWYRPKOLKDGTTWDSKETDMRPIILMVWMPNLTQAYVLYNMKQGNLLPSALPFF 348
Db 296 ANTWYRPKELKDGKWTASTENDKRPLLTVVWMPSKAIQASLYNMKQEGGLGTNTQYTFSS 355
QY 349 NADA--DPAELNHYSEIVQONIEKRISGTNDWLTMLHDFVYNNPMNKKDSNVNRSFG 406
Db 356 SSQTQMDQAAL-----VQKRIGRIARBSNTDWLTATIKNEVKTQPGWNSSTENLD--NN 409
QY 407 IQQGGFLAYKENSIDLTPYANSYRLLGRPIN-----TKDQYRGOEFLLANDIDN 457
Db 410 DHLQGGALLYNDSRTSHANSYRLLNRPTTQTKHNPKYTKDTSNGGFEFLLANDIDN 469
QY 458 SNPVQAEOLNWLNYLLNFGTITANNDQANFSRVYDAPNDIDALMNIQAQDYENRAGVM 517
Db 470 SNPAVQAEOLNWLHLYMNIIGTITGSGEDENFQVYDADVNDVNDLLQIASDYFKAQGA 529
QY 518 D-SDAVSNKHINILEDNMHADPEYFNKIGNPOLMTDDTIKNSLNHG--LSDATNRMGLDAI 575
Db 530 DQSQDQAIKHLISILEANSHNDAYNEDTKGAOLPMDDPHMLALVYLLAPIGNRSGVPEL 589
QY 576 VHQSLADRENNSTENVIPNYSFVRAHDNNSQDQIQNALRD-----VTGKDYHTTFEDE 630
Db 590 ISNSLDRSESGKSKRMANYAFVRAHDSEVOSIIGQIIKNEINPOSTG---NTTFLDEM 646
QY 631 OKGIDAYIODNSTVKKYLYNIPASVAILLTNKNDRIPRYVYGDLYTDGGOYMEHQTRY 690
Db 647 KKAFFIYNKDMRSANKQYQYINPISAYALMLHKDVPVYIGDMITDDGQYMAOKSPIY 706
QY 691 DLTNLLKSRVYVAGSQSMQTSVG--GNNN-----ILTSVRYGKGAMTATDTGDETR 743
Db 707 DAETLLKGRIVAAAGQDMKVNYIGYGTNGWDAAAGVLTSVRYGFGANSASDTGTAEFR 766
QY 744 TQIGIVVYVSTNPLKLGVDKVVYLMHGAHKNOQYRAVLTITTDGVINVTSDGAP--VA 801
Db 767 NOGMVIVSNQPALRTSN--JTINNGAHRNQAYRPLLTNDGVATYLNDSANGIVK 824
QY 802 MTDENGDLYLSSHNLVYNGKEADTAVQYANPDVSGYLAWVPVPGASQNDQARTAPSTE 861
Db 825 YTDGNGNLFSANE-----IRGNPQVQGLAVVWPVPGASEQDVVRVAPKE 872
QY 862 KNSGSAYRTNAADFNSWIFEPASNFVYPTTKESERANVRIQANADFFASLGFTSEMAP 921
Db 873 KNSGLVYESNAALDSQVITYEGFSNFQDFVPNPSQYTNKKIAENANLPKSWGTSSEFAP 932
QY 922 QYNSSKDRFTLDSTIDNGVAFTRDYLGHSEPNKYCTDELRNAILQALHAKGLQWADWV 981
Db 933 QYVSSDDGSEFLDSVITQNGTAFTRIDIGMSKNKYGSLADLKAALKSLHAYGISAADWV 992
QY 982 PQIYVNLPGKEVATVTRVDDRGVNWKDAIINNLYVYVNT--IGGGEYQKKGAGFLDKLOK 1040
Db 993 PQIYVNLPGDEVVATRVNNYGETKDGALIDHSLYAAKVTFTGNDYQGYGAGFLDELAK 1052
QY 1041 LYPEITKQVSGVAIDPSQKITWSAKYFNGTNILHRSGVYVLKADGGQYVNLGTT--T 1099
Db 1053 LYPIEDRVYIISGTGKRTTDEKITKWSAKYMGNTILDRGSEYVLKNGLNGY--GTNGG 1110
QY 1100 KQPLTQLGEGKQKQNEGFVKGNDGN-----YFVDLAGNMMVKNFTIEYEDSVGNWY 1149
Db 1111 KYSLP-KYVG-SNQSTINGNDGSGKKEKELFVRYRYN-NGYAKNAFTRKNDGNVY 1167
QY 1150 FPDQDGKVENKHFVDVDSYGEKGTFFFLKNGVSFRGLVQTDNG--TYFEDNYGKVV-- 1205
Db 1168 YFDSGRMAVGEXTID---GKQ--YFELANGVQLRDGYRQNERGOVFFYDQNGVLSANG 1221
QY 1206 -----RQNTINAGAMIY--LDENGKLIKASYNSDAE 1235
Db 1222 KODPKPDNNNTSGRNFQVQIGNVWAYVDGNGKRYIGHQNINGQ 1266

RESULT 8
Q55264 PRELIMINARY; PRT: 1449 AA.
ID Q55264
AC Q55264;


```

DT 01-NOV-1996 (T-EMBLrel. 01, Created)
DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (T-EMBLrel. 17, Last annotation update)
DE GLUCOSYLTRANSFERASE PRECURSOR.
GN GN
OS Streptococcus salivarius.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1304;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95122197; PubMed=7822030;
RA Simpson C.L., Giffard P.M., Jacques N.A.;
RT Streptococcus salivarius ATCC 25975 possesses at least two genes
RT coding for primer-independent glucosyltransferases.*;
RL Infect. Immun. 63:609-621(1995).
DR EMBL; L35495; AAC41412.1; -.
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 8.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Signal; Transferase.
KW SIGNAL 1 35 POTENTIAL.
FT CHAIN 36 1449 GLUCOSYLTRANSFERASE.
FT CHAIN 36 1449
SQ SEQUENCE 1449 AA; 159984 MW; DP62807306E86A46 CRC64;

Query Match 35.7%; Score 2371; DB 2; Length 1449;
best Local Similarity 42.4%; Pred. No. 6.1e-101;
Matches 553; Conservative 172; Mismatches 416; Indels 164; Gaps 40;

QY 5 ARPDVAAYPSLYNSAVSGD--TTIKLTNDQYQALNGQLQVLLRFSKAADGNPSGDNTV 62
DB 52 AIPDITVDTGTVSDTTAAQDPTTAAATND---VATQAFPTATEDLTDTT-----NTV 104
QY 63 TDQFSKNYATGGH-----EDYKVNQNGYVEFGSHATNQSDKDSQWIIVLVNGKEVKR 117
DB 105 AANAVDTVATGTDRAATTNDTATNDTAVD-----TTNNNTTDT---TTVDRAATTE 156
QY 118 QLVNDTEGAAGFNNDVYKVPAINSSHSFGQIITLPTVKNNVQLVHFRSNDVKT 177
DB 157 KRATGARGGTGGER--ATPVNGTNNANNT-----VTVVNDLPA-----TNNVVT 201
QY 178 GEGNYVDFWSELMPVKDFQKNGPLKQFGLQTNGQYVIDPTGTPGPRKNFLQSGNNW 237
DB 202 -----DGPSH---IKTLNGQYIYE-DDGTRIRYVLERIGGS 235
QY 238 IYFSDTGVGTNALELQPAK--GTVSSNEQYR-----NG--NAAYSYDDKSIENVNGYLT 288
DB 236 QYFNAETGELSNQKEYRFDKNGGTGSGADSTNTNTVNGDKNAPFYGTTDKDIELVDGYFT 295
QY 289 ADTWYRKQILKDTTWTDSKETDMRPIWVWPNTLTQAYLYLNKYNKQHGNLLPSALPFF 348
DB 296 ANTWYRKELKDKGKWTASTENDKRPLLTVMWPSKAQSYLYNKEQSLGNTQYTF 355
QY 349 NADA--DPAELNHYSEIVQVQNIKRISETGNTDLRLTLMHDFVNNPNMKNKDSNVNFSG 406
DB 356 SSQQTMDQAAL-----VQKEIERIAREGNTDLRTIANKFYATQGHNSTSNLD--NN 409
QY 407 IQQGGFLKYENSDLTPYANSYKLLGRMPTN-----IKQFYRGQEFFLLANDIDN 457
DB 410 DHLQGGALLYNNDSTSHANSYKLLNRTPTSPQKHNPKYTKDTNNGGFEFLANDIDN 469
QY 458 SNPVVQAPQLNWLXYLLNFGTITANNQAFDSVRVDPONIDADIAMNIADYFNAYGM 517
DB 470 SNPAVQAPQLNWLXYLLNFGTITANNQAFDSVRVDPONIDADIAMNIADYFNAYGM 529
QY 518 D-SDAVSNKHNILEEDNNDHAPYFNKIGNPQLTMDTINKSNHIG--LSDATNRWGLDAI 575
DB 530 DQSQDQATKHLISLEASHNDAYNEDTKAQALPMDPPHMLALVYLLIRIGNSGYEPL 589
QY 576 VHQSLADRENSTENVIPNYSFVRHNNNSQDQIONAIRD-----VTGKDHPHFTFEDE 630
DB 576 VHQSLADRENSTENVIPNYSFVRHNNNSQDQIONAIRD-----VTGKDHPHFTFEDE 630

```

```

DB 590 ISNSLNDRESGKSKRMANYAFVRAHDSVQSIIGQIIKNEINPQSIG---NTFTLOEM 646
QY 631 QKGIDAYIODNSTVKKYNLYNIPASYAILLTKNDTIPRVYVYGDLYTDGQYMEHQTRY 690
DB 647 KKAFTYKNDMSANKQYTOYNIPSAVALMLTKDTPRVYVYGDYTDGQYMAOKSPY 706
QY 691 DLTNLILKSRVYVAGGQSQMPTMSVG--GNNN-----ILTSVYKGMAMTATDTDETR 743
DB 707 DAETLLKGRIRIYAAGGQDMKYNIGYGTNGNDAGVLTSVRYGTGANSASDTQTAETR 766
QY 744 TQIGVGVVNTNKLGVNDKVVLMHGAARKNOQYRAAVALTTTDDGVINVTSDQAP--VA 801
DB 767 NQGMVIVSNQFALRTSN--LTINMGAAHRNQARPLLTNDGVAYLYLNDSDANGIVK 824
QY 802 MTDENGDLTSSHNLLVYNGKEADTAQVGYANPDVSGYLAVVVPVGCASDNQDARTAPSTE 861
DB 825 YTDGNGNLATFSANE-----IRGIRNPQVDGYLAVVVPVGCASEMDQVRVAPSK 872
QY 862 KXSGNSAYTNAAFDSNVIFEAFSNFVYPTKESERANVRIAGNADFFASLGTTFEFAP 921
DB 873 KXSSGLVYESNAALDSQVIYEGFSNFQDFVQNPQSYTNKKIAENANLFSKSGITSEFAP 932
QY 922 QYNSKDRFTFLDSTIDNGYAFTRDYDLGHSEPNKYGTDRDLRNLALQALHKAQIYQWADWV 981
DB 933 QYVSSDGSFSDSVIQNGYAFTRDYDIGHSKDKYGLADLKAALSKLHVAISAIADWV 992
QY 982 PQIYNLPKEVATVTRDGRNVWKAAILNNLYVYNT--IGGEYQKKGAGFLDKLQK 1040
DB 993 PQIYNLPDGVATVTRVYNGYGETKDGAIIHSLYAAKTRTENDYQKYGAGFLDELKR 1052
QY 1041 LYPEITFKQVSGVAIDFSQKITMSAKYFNGTNTLHRSGYVLKADGGQYVNLGTT--T 1099
DB 1053 LYQIIFDRVQISTGKRMKTTDEKITQMSAKYMTNLDGSEYVLKANGLYY--GTING 1110
QY 1100 KQFLPQLTGEKKQSGNEGFVKGNDGN-----YFYDLAGNKNVKNFTIEDSVGNWY 1149
DB 1111 KVSLP-KVVG-SNOSTNGDNONGSGSKPEKRLFSVRYRYN-NGQYAKNAFLKNDGNYV 1167
QY 1150 FPDQDKKVENKHFVDVDSGEKGTFFFLKNGVSFRGLVQTDNG--TYFYDNKGM---- 1204
DB 1168 YFDSGRMAVGKTIID---GKO--YFFLANGVQLRDGYRQNRGQVFFYDQNGVLNANG 1221
QY 1205 -----VRNQITNAGAMIYT-LDENGKLKASVNSDAE 1235
DB 1222 KQDPKPDNNNSGRNQFVQIGNVWYAYDNGKRVTHGQNINGQ 1266

RESULT 9
O69397
AC O69397 PRELIMINARY; PRT; 1455 AA.
DT 01-AUG-1998 (T-EMBLrel. 07, Created)
DT 01-AUG-1998 (T-EMBLrel. 07, Last sequence update)
DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)
DE GLUCOSYLTRANSFERASE-SI.
GN GN
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=WT4467;
RX MEDLINE=98231643; PubMed=9570124;
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,
RA Kilmura S., Hamada S.;
RT *Molecular analyses of glucosyltransferase genes among strains of
RT Streptococcus mutans.*;
RL FEMS Microbiol. Lett. 161:331-336(1998).
DR EMBL; D89978; BAA26120.1; -.
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 9.

```


DR pfam: pf02324; Glyco_hydro_70; 1.
KW transferase.
SQ SEQUENCE 1455 AA; 162914 MW; 1263427BP24E9E1 CRC64;

Query Match 35.3%; Score 2346.5; DB 2; Length 1455;
Best Local Similarity 41.0%; Pred. No. 8.2e-100;
Matches 536; Conservative 188; Mismatches 397; Indels 187; Gaps 37;

QY 19 SAVSGFDTTIKLTNDQYQALNGQLVLLRFSKA-----ADGNPSG---DN 60
Db TSLSGLVKADSTDRQQAATVSQASLVTSSAAKETLTATDTSTATSQTATVTDN 88
QY 61 TYTDQFSKNVATGGNEDVYKNGQVSGHATNQSNDKDSQWIIIVLNGKEVKRLV 120
Db VSTNQSTNTTANTANFVKPTTSS-----QAKTDSK-----IATSKAVNLT- 135
QY 121 NDTKEGAAGFNRRNDYKVNPAITENSSMSGFGQIITLPYTVKNENYVQLVHRFSNDYKTGG 180
Db 136 -----ATG-----KFPVANNNTAH-----PKIVTD----- 155
QY 181 NYDWFSELMVPKDFQKNGPLQK-----FGLQTINGQOYYIDPTTQOP 225
Db 156 -----KIVPIKPKI-----GKLQKPSLSQSDIAALGNVKNIRKVGKYYY-KEDGTL 203
QY 226 RKNFLQSCNNWNYFSDTGVGTNALELQFAKGTYSNE--QVRNGNAAVSYDDKSTEN 282
Db 204 QXNYALNINCKTFFDEYGAISNNVLPK--KGNITNDNTNSFAQYQVYSTDVANFEH 261
QY 283 VNGYLTADWPYRPKILKGGTWTWDSKETDMPITLVMVWPNLTQAYLYNVKOHGNLIP 342
Db 262 VDHLYTASRWPKYILKDKGTQSTQTEKDRPPLMTWPDQETQKQVNYMNAQLGIHQ 321
QY 343 SALPENADAPAEALNHVSEIVQONIEKRISTGTWDLRLHDFVTVNPNWKNKDSNV 402
Db 322 T-----YNTATSPQLNLAQTITQIEKIKTAETKNTMLRQTIAPFVKTQSAWNSDSKPK 377
QY 403 NFGIQFGQFGLKYENSIDTPYANSDYRLGLRMPIN-----IKDQTVRGQBFLLAN 453
Db 378 FDDHLQ-KGALLYSNSKLTQANSNYRLNTPNTQKGDPRYATDRTGGYEFLLAN 436
QY 454 DTDNSNPPVQABQLNWLHYLLNFGIT(TANNQANSDSVRVDAPODNIDALMNAIDYFNA 513
Db 437 DYNDSNPPVQABQLNWLHFLMFGNIYANDPDANFDSIRVDAVNDVADLQIAGDYLKA 496
QY 514 AXGM-DSDAVSNKHINILEDWNHDAPEYFNKIGNQLQMDDTIKKSLNHLSDAN-RWG 571
Db 497 AKGIHKNKAANDHLSILEAWSYNDTPYLLHDDGNNINMDNRLRLSLYSLSAKPLNRSG 556
QY 572 LDAIVHQLADRENNSTENVIPNYSFVRAHDNNSQDQTONAIR-----DVTGKDYHFT 626
Db 557 MNPLTNSLVNRTDQNAETAAPVPSFIRAHDSVQDLIRDIKAEINPNVVG---YSFT 613
QY 627 FDEQKGDAYIQDQNSVTKYKLNLYNIPASVAILTNTKOTIPRVYVGLDYGQGYMEHQ 686
Db 614 MBEIKKAFETYNKULLATEKKYTHYNTALSVALLTNKSVPVRYVGMFTDDGQYMAHK 673
QY 687 TRYDYLNLKSRVYKYGQSMQTSVGGNNNTLTQVRYGKGAWTADTGTDTTRQG 746
Db 674 TINYPAIETLLAKRIKYSGGQMRQVY-GNSEITSVRYGKALKATDGTDRFTTSG 732
QY 747 IGVVYSNTPNLKLVNDKYLHMGAAHKNQOYRAAVLTATDGVINYTSDQGA--VAMTD 804
Db 733 VAVIEGNPSLRLKASDRVYVNGAAHKNQAYRPLLLTDDNGIKAYSDQRAAGLYRTN 792
QY 805 ENGDIYLSHNLVNGKEADTAVQGYANPDVSGVYAVVPGASDNQDAPAPSTERN 864
Db 793 DRGELITFA-----AD--IKGYANPQVSGVGLVWVPVGAADQDVRVAASAPST 840
QY 865 GNSAYRTAAEDSNVIEAFPSNFVYTPKESERANVYTAQADFPASLGFSFENAPQYN 924
Db 841 DKSQVHQNALDQWFEFGFSNQAFATKKEEYTNVYAKNVDRFAEWGVTDFEMAPQVY 900

QY 925 SSKDRTFLDSTIDNGYAFTRDYDLGMSBPNKYGTDEDLRNAIQALHKAQLQVMAWVDPQ 984
Db 901 SSTDGSFLDSVIQNGYAFTRDYDLGMSBPNKYGTDEDLRNAIQALHKAQLQVMAWVDPQ 960
QY 985 IYNFGKGVATVTRVDRDGNVWKDAIINNLLVYVN-TIGGGEYOKKYGGAFDLKLOKLYP 1043
Db 961 MYALPEKEVVTATRVKGTVPAGSOIKNTLYVVDGKSSGKQQAQKAGAFLEELQAKYP 1020
QY 1044 EIFTKKQVSTGVAIDPSOKITWESAKYFNGNIIHRGSGYVVKADGGQYVNLGTTTKQFL 1103
Db 1021 ELFAKQISTGVPMDDPSVKIKQWSAKYFNGNIIHRGSGYVVKADGGQYVNLGTTTKQFL 1080
QY 1104 PIQLTGKQKQNEGKFGK--NDG-NYIYDLAGNMVKNFFIEDSVG-NWYFFDQDGKAVE 1159
Db 1081 PRSLV-NPNHGTSSSVTLGVFDGKGYVYSTSGNAQNAFI--SLGNWYFYDNGGNYVT 1137
QY 1160 NKHFVDVDSYGEKGYTFYFKNGVSRFGGLVQT-----DNGTY-----Y 1197
Db 1138 GAQISN-----GANYFSLNGIQLRNAIYDNGKNVLSYNGDGRYENGYLFGQWRY 1191
QY 1198 FDTGKGVNRTINAGMIYTLDENGKLIKASYNDAEYPTSTDVGM 1245
Db 1192 FCN-GIMAVGLTRVHGAVQY-FDASG-----FOAKGQFITAD-GKL 1230

RESULT 10
O69388 PRELIMINARY; PRT; 1455 AA.
AC O69388;
DT 01-AUG-1998 (TREMBlrel. 07, Created)
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE GLUCOSYLTRANSFERASE-SI.
GN GTFC.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID:1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MT4239;
RA MEDLINE-98231643; PubMed-9570124;
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,
RA Kinura S., Hamada S.;
RT *Molecular analyses of glucosyltransferase genes among strains of
RT Streptococcus mutans.;
RL FEMS Microbiol. Lett. 161:331-336(1998).
DR EXBL; D88658; BAA26110.1; -;
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 10.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW transferase.
SQ SEQUENCE 1455 AA; 163046 MW; 6D90A4978D35DD82 CRC64;

Query Match 35.3%; Score 2343.5; DB 2; Length 1455;
Best Local Similarity 41.1%; Pred. No. 1.1e-99;
Matches 534; Conservative 187; Mismatches 409; Indels 169; Gaps 35;

QY 19 SAVSGFDTTIKLTNDQYQALNGQLVLLRFSKAADGNPSGDNT-----YTDQFSKN 69
Db 29 TSLSGLVKADSTDRQQAATVSQASLVTSSAAKETLTATDTSTATSQTATVTDN 88
QY 70 YATGGNPDYKVGNGOVFEFGHATNQSNDKDSQWIIIVLNGKEVKRLVNDTKEGAAG 129
Db 89 VSTNQSTNTTANTANFVKPTTSS-----QAKTDSK-----IATSKAVNLT-----ATG 138
QY 130 ENRDYKVNPAIENSSMSGFGQIITLPYTVKNENYVQLVHRFSNDYKTGGYVDFWSEL 189
Db 139 -----KFPVANNNTAH-----PKIVTD-----KI 157
QY 190 MPVKDSFQKNGPLQK-----FGLQTINGQOYYIDPTTQPRKNFLQSG 234

528 NILEDDNHADPEYKNGKIGNPQMTMDTINKSLNHLGLSDATN--RMGLDAIVHQSADRENN 566
512 SILEAWSYNDTPYLHDDGDNWMLNMDRLSLSLSLAKPLNORSCHNPITTSNLVNRD 571
587 STENVVPIPNYSFVRADHNSDOQIQNAIR-----DVTGKDYHTFTFEDEQKIDAYIQDQ 641
572 NAEATAVPSYSFIRAHDSVQDLIRDIKAENPNVVG---YSFTMEELKKAFAEYLNKDL 628
642 NSTVKKYNYLNPASVAILLTKNDKIDIPRYVYDGLTDDGGYMEHQTFRYDTLNLKSRV 701
629 LATEKKYTHYNTALSYALLTNKSSVPRVYVYDGMFTDDGGYMAHKTINYEATETLLKARI 688
702 KYVAGQSMOTWSVGGNNILTSVRGKGMATATDGTDETRTQIGVYVSTNPNLKGV 761
689 KYVSGQAMRNQV--GNSLIITSVRIGKALKATDGRITTSVGVVIEGNNPRLRKA 747
762 NDKVYLHMGAAHKNQOYRAAVLTITDGVINYSQDQAP--VAMTDENGDLYLSSHLVYN 819
748 SDRVVVMGAAHKNQOYRPLLTITDNGIKAYHSDQEAAGLVRYTNDRGELIFTA----- 801
820 GKEADTAVQGANPDVSGYLAWVPVVGASDQDARTAPSTKNSGNSAYRTNAAFDSNV 879
802 ----AD--TKGANPQVSGYLGWVPVVGAAADQDVRVAASTAPSTDKSVHQAALDSRV 855
880 IFEAFSNFYVTTKESERANVRIAGNADFFASLGFTSPREMAPOYNSSKDRTFDSTIDNG 939
856 MFEFGSNFOAFATKEEYTNVIAKNVFAEWGVTDFEMAPQYVSSSTGSLDSVIONG 915
940 YAFDORYDGLGMEPNKYGDEBLRNAIQALHAKAGLYMADWVPDQIYNLPKGEVATVTRV 999
916 YAFDORYDLGSKPNKYGTDADLVAIKALHSGKIKVMADWVPDQWYALPEKEVTVATRV 975
1000 DRRGNVRKDAITNNLYVN--TIGGEGYQKYGGAFLDKLQKLYPEIFTKKQVSTGYAID 1058
976 DKYGPVAGSQIKNTLYVVDGSSKQDQAKYGGAPLEQLQAKYPELFARKQISTGVPM 1035
1059 PSQKTEMSAKYFNENILHRGSGYVLKADGGQYVNLGTTTKQFPIQITGKKGNGEFG 1118
1036 PSVKIKOWSAKYFNENILHRGSGYVLKADGGQYVNLGTTTKQFPIQITGKKGNGEFG 1094
1119 VKG--NDG--NYFYDLAGNKNWTFIEDSVG--NWTFDODGKAKENKHFVDVDSYGEKGT 1174
1095 VTGLVFDGKGYVYVYSGNAKNAFI--SLGNWYVYFDMNGYVAVGAQSN-----GAN 1146
1175 YFLKNGVSGFRGLVQT-----DNQTY-----YFDNGYKAVRQNTINA 1212
1147 YFLSNGIQLRAIYDNGKNVLSYVGNOCRRYENGYYLFGQQRIFQNG--GIMAVGILRVH 1205
1213 GAMIYTLDENGKLIRASVNSDAEYPTSTDVGRH 1245
1206 GAIQY--FQASG-----FQAKGQFITTAD--GKL 1230
RESULT 12
069385 PRELIMINARY; PRT; 1390 AA.
AC 069385;
DT 01-AUG-1998 (Tremblrel. 07, Created)
DT 01-AUG-1998 (Tremblrel. 07, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE GLUCOSYLTRANSFERASE-SI.
GN GTFP.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus
OX NCBI_TaxID:1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-WI4245;
RX MEDLINE=98231643; PubMed=9570124;
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,
Kimura S., Hamada S.;

*Molecular analyses of glucosyltransferase genes among strains of
Streptococcus mutans.;
RL FEMS Microbiol. Lett. 161:331-336(1998).
DR EMBL; D88655; BAA26106.1; -;
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 7.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW transferase.
SQ SEQUENCE 1390 AA; 155375 MW; 8847E4956EF05E9F CRC64;
Query Match 35.3%; Score 2341.5; DB 2; Length 1390;
Best Local Similarity 41.4%; Pred. No. 1.3e-99;
Matches 535; Conservative 187; Mismatches 414; Indels 157; Gaps 35;
QY 19 SAVSFTDTIKLTNDQYQALNGQLQVLRFSKA-----ADGNPSG---DN 60
DB 29 TSLGSLVKAADSTDDROQAVTESQASLVTTSAAKETLTATDTSTATSATSQPTATVTDN 88
QY 61 TVTDFSKNYATTTGNGFDYVYVNGNQVEFSGWHTNQSNDKDSQWIIIVLVNGKEVKQLV 120
DB 89 VSTTNQSTNTTANTANFDVKPTTSE-----QSKTDNSDK-----IATSKAVNRLT- 135
QY 121 NDTKEGAGFERNVDYKYNPALENSMSFGOIIITLPYVANKENVQLVHRFSNDVKYEGE 180
DB 136 -----ATG-----KEVPANNNTAHSRTVTDKIVPIKPKIGLKQPPSSLSODDIAALG 182
QY 181 NYVDWFSELMKPKVDSFKNGPLAQFGQTINGQQYYIDPTTGPGRKNFLQSGNNWYIF 240
DB 183 N-----VKN-----TRKNGKYVY--KEDGTLQKNVALNKGKTFFF 218
QY 241 DSDTGVGTNALELQPAKGTVSNE---QYRNGNAAYSDDKSIEENVNGYLPADTWYRPKQ 297
DB 219 DETGALSNNILPSK--KGNITNNDNTNSFAQYNOVYSTDAANFEHVDHYLTAESWYRPKY 276
QY 298 ILKDGTTWTDSKETDMRPILMVMWNTLTQAYLYNMQHGNLLPSALPFPNADAPAE 357
DB 277 ILKDGTTWTQTEKDFRLMTWHPDQETQYQYVYVMAQGIHQT-----YNTATSPLOL 332
QY 358 NHYSTVOONTEKRISGTGNTDMLRTLMHDFVTNPNMKNDSNVNFSGIGQGFGLKYE 417
DB 333 NLAQTIQTKIEEKITAENKTNLWRTISAFVKTSANNSDSEKPFDDHLQ--KGALYLSN 391
QY 418 NSDLTPYANSYRLLGRPIN-----TKDQYRQEFLLANDINDNSNPVQAEOLN 468
DB 392 NSKLTQANSYRILNRTPTNQTKKOPRYTADRTIGGYEFLANDVNSNPVQAEOLN 451
QY 469 WLYYLLNFGTITANDQANFDSVRVDAPDNIDADLNLNIAOYFENAAVGM--DSDAVSNKHI 527
DB 452 WLHFLMNEGNIYANDPDANFDSIRVDVNDVADLLQTAGDYLRKAAGIKHKNDAANDHL 511
QY 528 NILEDDNHADPEYKNGKIGNPQMTMDTINKSLNHLGLSDATN--RMGLDAIVHQSADRENN 586
DB 512 SILEAWSYNDTPYLHDDGDNWMLNMDRLSLSLSLAKPLNORSCHNPITTSNLVNRD 571
QY 587 STENVVPIPNYSFVRADHNSDOQIQNAIR-----DVTGKDYHTFTFEDEQKIDAYIQDQ 641
DB 572 NAEATAVPSYSFIRAHDSVQDLIRDIKAENPNVVG---YSFTMEELKKAFAEYLNKDL 628
QY 642 NSTVKKYNYLNPASVAILLTKNDKIDIPRYVYDGLTDDGGYMEHQTFRYDTLNLKSRV 701
DB 629 LATEKKYTHYNTALSYALLTNKSSVPRVYVYDGMFTDDGGYMAHKTINYEATETLLKARI 688
QY 702 KYVAGQSMOTWSVGGNNILTSVRGKGMATATDGTDETRTQIGVYVSTNPNLKGV 761
DB 689 KYVSGQAMRNQV--GNSLIITSVRIGKALKATDGRITTSVGVVIEGNNPRLRKA 747
QY 762 NDKVYLHMGAAHKNQOYRAAVLTITDGVINYSQDQAP--VAMTDENGDLYLSSHLVYN 819
DB 748 SDRVVVMGAAHKNQOYRPLLTITDNGIKAYHSDQEAAGLVRYTNDRGELIFTA----- 801
QY 820 GKEADTAVQGANPDVSGYLAWVPVVGASDQDARTAPSTKNSGNSAYRTNAAFDSNV 879

```
Db 802 ----AD--IKGYANPQVSGYLGWVPVGAADQDVRAAASPATDCKSVHQNAALDSRV 855
Qy 880 IFEAFSPNFVPTKESERANRYKIAQNAOFFASLGFTSEMAPOYNSSKDRDTFLDSTIDNG 939
Db 856 MFEFGSNFQAPATKKEEYTNVIAKNVXKFAEWGVTDEMAPOYVSTSDGSFLDSVQNG 915
Qy 940 YAFDRYDLGNSEPNKYGTDBDLRNAIQALHAGLOVMAWVPDQIYNLPKGEVATVTRV 999
Db 916 YAFDRYDLGISKPNKYGTADLVKAIALHSGIKVMADWVPDQVYALPEKEVYVTRV 975
Qy 1000 DDRGNVWKDALINNNLYVN--TIGGEYQKYGAFDLKLOKLPELFTKKQVSTGVAID 1058
Db 976 DKYGTVPAGSQIKNTLYVVDKSSGQQAQYGGAFLELQAKYPELFARKQISTGVPM 1035
Qy 1059 PSOKITEWSAKYNGTNILHRGSGYVLKADGGQYVNLGTTTKQFLPQLTGEKKQNEGF 1118
Db 1036 PSVKIKQWSAKYNGTNILHRGSGYVLKADGGQYVNLGTTTKQFLPQLTGEKKQNEGF 1094
Qy 1119 VKG--NDG--NYEFDLAGNMVKNFTIEDSVG--NWYFFDQDKVMKNKHFVDVDSYGEKGT 1174
Db 1095 VTGLVFDGKGYYVYSTSGNAKNAFI--SLGNWYFDNGYVMTGAQSIN-----GAN 1146
Qy 1175 YFELKNGYFSGELVQT-----DNQTY-----YEDNYGKVMNQVINA 1212
Db 1147 YFELKNGYFSGELVQT-----DNQTY-----YEDNYGKVMNQVINA 1205
Qy 1213 GAMIYTLDENGKLKASYNDAEYPTSTDVGKM 1245
Db 1206 GAVQY-FDASG-----FOAKGQFTTTAD-GKL 1230

RESULT 13
O69391 PRELIMINARY; PRT: 1455 AA.
AC O69391;
DT 01-AUG-1998 (TReMBLrel. 07, Created)
DT 01-AUG-1998 (TReMBLrel. 07, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE GLOUCOSYLTRANSFERASE-SI.
GN GTEC.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus
OX NCBI_TaxID:1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-W4251;
RX MEDLINE-98231643; PubMed-9570124;
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,
RA Kinura S., Hamada S.;
RT "Molecular analyses of glucosyltransferase genes among strains of
RT Streptococcus mutans.";
RL FEMS Microbiol. Lett. 161:331-336(1998).
DR EMBL; D88661; BAA26114.1;
DR FMS Microbiol. Lett. 161:331-336(1998).
DR InterPro; IPR02479; CW binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 9.
DR Pfam; PF02324; Glyco_hydro_70; 1.
RW transferase.
SQ SEQUENCE 1455 AA; 162804 MW; 683A359D873E9E1A CRC64;
```

```
Query Match 35.2%; Score 2340.5; DB 2; Length 1455;
Best Local Similarity 41.4%; Pred. No. 1.5e-99;
Matches 535; Conservative 186; Mismatches 415; Indels 157; Gaps 35;
Qy 19 SAVSGFDITIKITNDQYALNQLQVLLRFSA-----ADGNPSSG---DN 60
Db 29 TSLSGSLWKADTDURQQAQVATESQASLTYSAAKETLTATDTSTATSTQPTATVTDN 88
Qy 61 -TVDQFSKNYATGCGNFQVKNVQGVSGWHAHQNSNDKDSQWIIIVLVNGKEYKQLV 120
```

```
Db 89 VSTNOSTNTANTANFVDPKTTTSE-----OSKTDNSDK-----IATSKAVNRLT- 135
Qy 121 NDAKEGAAGFNRNDVYKVNPAIENSSMSGFGIITLPVTVKNENVOLVHRFSNDVKTGEG 180
Db 136 -----ATG-----KVFANNNTAHSRTVTKIPIKPKIGLKQPSLSQDDIAALG 182
Qy 181 NYVDFWSELMPVKSQKNGPKLQFGLOPIINGQQYIDPTTGOPRKNFLLQSGNNHIYF 240
Db 183 N-----VKN-----IRKVNKYYVY-KEDGTLOKNYALNNGKTTFF 218
Qy 241 DSDTGYGTNALEQFAKGTYSNE--QYRNGNAAYSYDDKSIENYNGYLTADTWPRKQ 297
Db 219 DETGALSNNLPSK--KGNITNNDNTNSEAYNOYNTDAANFEHVDHYLTABSWRPXY 276
Qy 298 ILKDGTTWDSKSTDMRPILMVWVWPNFLTQAYLTNYMKQHGNNLLPSALPFFNADAPAE 357
Db 277 ILKDGKTWOSTEKDPRLLMTWVQDETQORQVYVYNAAGLGHQ?-----YNTATSELQL 332
Qy 358 NHYSEIVQOQMEKRISTGNTDMLRTLMDHDFVTNNPMWKNKDSNVNPSFGIQFGGFLKY 417
Db 333 NLAQQTQTKIEKITAENKTNMLROTISAFVKTQSAWNSDSEKPPDDHLQ-KGALLYSN 391
Qy 418 NSDITPYANDYRLGEMPIN-----IKDQTYGQBEFLANDIDNSNPVYQAEQLN 468
Db 392 NSKLTQANSNRYRLNRPNTNQTKGKDPRTADNTIGGYEFLANDVDNSNPVYQAEQLN 451
Qy 469 WLYYLLNFGTITANNQANFDSVRYDAPDNIDADLNMIAQDYFNAAYGM-DSDAVSNKHI 527
Db 452 WLHFLMFGNIYANDPDANFDSINVDVNDVADLLQIAGDYLKAAKGINKDKAANDHL 511
Qy 528 NILEDWHAADPEYFNKIGNPQLTMDDTIKNSLHGLSDATN-RWGLDAIVHQSLADRENN 586
Db 512 SILEAWSDNDTPYLHDDGDNNIMNDKRLSLFLSLAKPLNQRSGMNPILITNSLVNRD 571
Qy 587 STENVVTPNYSFVRAHDNNSODQIONAIR-----DVTGKDYHFTFEDDQKGDYATDQ 641
Db 572 NAETAAVPSYFIRAHSEVQDLRIDIKAEINPVVG---SYFTEEEIKKAEIYNKOL 628
Qy 642 NSTVKYNLYNIPASYAILTNKDTIPRVYGDLYTDGQYMEHQTYRYDYLTLMLKSRV 701
Db 629 LATEKTYHTYNTALSYALLTNKSVPRVYVGMFTDDGYMAHKTINYEAETLLKARI 688
Qy 702 KYVAGGSHQMTMSVGGANNILTSVRYGKGAMTATDCTDFTQGTGIVGVSNPFLKLV 761
Db 689 KYVSGGAMRNQOY-GNSEIITSVRYGKGAKATDGTDRTRTSGVAVIEGNNP SLRKA 747
Qy 762 NDKVYLHNGAAHKNQYRAAVLATTDGVIYNTSDQAP--VAMTDENGDLVLSHNLVYN 819
Db 748 SDRVYVNGAAHKNQYRPLLLTDNGIKAYHSQAEAGLVRYTNDRGELIFTR----- 801
Qy 820 GKEADTAVQYANPDVSGYLAWVPVVGASNDODARTAPSTEKNKSGNSAYRTNAFDSNY 879
Db 802 ---AD--IKGYANPQVSGYLGWVPVGAADQDVRAA--STAPSDGKSVHQNAALDSRV 855
Qy 880 IFEAFSPNFVPTKESERANRYKIAQNAOFFASLGFTSEMAPOYNSSKDRDTFLDSTIDNG 939
Db 856 MFEFGSNFQAPATKKEEYTNVIAKNVXKFAEWGVTDEMAPOYVSTSDGSFLDSVQNG 915
Qy 940 YAFDRYDLGNSEPNKYGTDBDLRNAIQALHAGLOVMAWVPDQIYNLPKGEVATVTRV 999
Db 916 YAFDRYDLGISKPNKYGTADLVKAIALHSGIKVMADWVPDQVYALPEKEVYVTRV 975
Qy 1000 DDRGNVWKDALINNNLYVN--TIGGEYQKYGAFDLKLOKLPELFTKKQVSTGVAID 1058
Db 976 DKYGTVPAGSQIKNTLYVVDKSSGQQAQYGGAFLELQAKYPELFARKQISTGVPM 1035
Qy 1059 PSOKITEWSAKYNGTNILHRGSGYVLKADGGQYVNLGTTTKQFLPQLTGEKKQNEGF 1118
Db 1036 PSVKIKQWSAKYNGTNILHRGSGYVLKADGGQYVNLGTTTKQFLPQLTGEKKQNEGF 1094
Qy 1119 VKG--NDG--NYEFDLAGNMVKNFTIEDSVG--NWYFFDQDKVMKNKHFVDVDSYGEKGT 1174
Db 1095 VTGLVFDGKGYYVYSTSGNAKNAFI--SLGNWYFDNGYVMTGAQSIN-----GAN 1146
```

```

QY 1175 YFELKNGVSFRGLVQF-----DNGTY-----YFDNYGKVRNOTINA 1212
DB 1147 YIFLNGQLRMAITDNGKVLVSYYGNGRYYENGYIIFGQWRYFQNGINAVGLRIH 1205
QY 1213 GAMIYTLDELNGKLIKASYNDAEYPTSTDVGKM 1245
DB 1206 GAVQY-FDASG-----FQAKGQFITTAD-GKL 1230

RESULT 14
Q9LCH3
ID Q9LCH3 PRELIMINARY; PRT; 1575 AA.
AC Q9LCH3
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE GLUCOSYLTRANSFERASE.
GN GTRF.
OS Streptococcus oralis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1303;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=20231779; PubMed=10768934;
RA Fujiwara T., Hoshino T., Ooshima T., Sobue S., Hamada S.;
RT "Purification, characterization, and molecular analysis of the gene
RL encoding glucosyltransferase from Streptococcus oralis.";
DR EMBL; AB025228; BAA95201.1;
DR InterPro; IPR002479; CW binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 17.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Transferase.
SQ SEQUENCE 1575 AA; 176792 MW; 772a26b4d7c2e543 CRC64;

Query Match 35.18; Score 2330; DB 2; Length 1575;
Best Local Similarity 42.69; Pred. No. 5.2e-99;
Matches 506; Conservative 175; Mismatches 372; Indels 134; Gaps 23;

QY 94 ATNQSNDKDSWII---YLVNGKEVKRLVND--TKGAAGFNNDYKVPALNDSMS 148
DB 161 AQDQEGDKREKTAVEDKIVANPKVAKKDLPEKSGQGAIA-ERMVADQAQFA 212
QY 149 GFQGIITLPVTYKNNVOLVHRFSNDYKTBEGNTYDWSSELMVPKDSFQKNGPLKQFGL 208
DB 213 -----PVNADHDDVLSH-----I 226
QY 209 QTINGQOYIIDPTGQPKNFLLQSGNNWIYFSDTGVGTNALELQFAKGVVSSNEQVRN 268
DB 227 KTIIDGKNTYVO-DGTGKKNFAVELNGILYEDATGALVDNEYQFOQGGTSSLNFEFSQ 285
QY 269 GNAAYSDDKSIENGVNGYLTADTYRPRKQIILKDGWTWDSKETMRPLMVMWPTLQQA 328
DB 286 KNAFGTITDKIEFDVGLTADSWYRPFILKDKTWTASTETDLRPLMAMWPDKRTQI 345
QY 329 YLYNMYKQGNLPLSALFFFNADAPAEINHYSEIVQONIEKRISSETGTDPLRLTMHDF 388
DB 346 NYLNTMNNQGG---LGAGAFENKVEQALLTASQSQVQRKIEKIGKEDTKWLTLMGAF 401
QY 389 VTNNPMWN--KDSNVNPSGTFQCGCFLLKYENSOLDTPYANSDYRLLRGRLN-----I 439
DB 402 VKTQNNHNIKESETGPKKDLHQLGALLYTNNEKSPHADSFRLLNRTPTTSQTGPKYF 461
QY 440 KDQYRGOEFLIANDIDNSNPVQAQELNMLYLLNFCITITANNDQANFDSVRYDADPNI 499
DB 462 IDKNGGYEFLANDFNSNFAVAQELNMLHYMWNFGSIVANDPTANFQDVRVDVNV 521
QY 500 DADLMLNTAQDYFNAAYGM-DSDAVSNKHINTLEOWNHADPEYFNKIGNPQLTMDITKNS 558

```

```

DB 522 NADLQIASDFKSKRYKVGSEEEAIRKLSILEAWSONDPDYNKDTGAQLAIDNKRLS 581
QY 559 LNHG-LSDATNRWGLDAIVHQSLADRENNSTENVIPNYSEVRAHDNNKSDQIQNALRD- 616
DB 582 LLYSEMRNLRSIRSGVEPTITNSLDRSSEKNGERMANYIFVRAHDSVQTVIADIIRN 641
QY 617 -VTGKDYHTFTFEDEQKIDAYIQDQNSTVKYKLYNIPASYAILLTKNKDTPRVYGD 675
DB 642 INPNTDGLTFTMDLQKAFKIYNEDMRKADKKYQFNIPTAHALMLSNKDSITRVYGD 701
QY 676 YTDGQYMEHOTRYVDTLNLLKSRVYVAGSQMOTMSVG-----GNNNLTISVRY 727
DB 702 YTDGQYMEKSPYHDAIDALIRAKIRYVAGGQMKYTYMGVPREADKWSYNGILTISVRY 761
QY 728 KGKAMATDTCGTDETGTGGIGVYVSNTPNLKGLYNDKVVILHMGAHKNQYRAAVITTD 787
DB 762 GTGANEATDECTAETRTQGMVIAASNPNLKLNEWDLQYNMGAAHAKNQYRPVLLITKD 821
QY 788 GVINYTSQGAAPVAM---TDNGDLYLSSHNLVYNGKEADTAVOGYANPDVSGYLAWY 844
DB 822 GISRYLADDEEYQSLWKKTANGILTDMND-----IAGYSNVQVSGYLAWV 869
QY 845 PVGASDNQDARTAPSTEKNNGNSAYRTNAAPDSNVIPAEFSNFYTTTKESERANVRIQ 904
DB 870 PVGAKADQDARTATSKKNASGVYESAALDSQLTYEGFSNFQDFATRDQYTNKYIAK 929
QY 905 NADFFASIGFTSFEMAFQYNSKDRITFLDSTIDNGYAFTDRYDLGMSEPNKYGTDEDLN 964
DB 930 NVNLFKEWGVTSFELPQYVSSQDGTFLDSIIQNGYAFEDRYDMAMSKNNKYGSLKDLN 989
QY 965 AIQALHAGLQVADWPDQIYNLPFGKAVATVTVDDRGVGNWKDAIINNLLYVYNT-IGG 1023
DB 990 ALRALHSVNIQAIADWPDQIYNLPFGKAVTATRVNNGYTYREGAEIKELIYVANSKTNE 1049
QY 1024 GEYQKYGGAFDLKOLYPEIFTKQVSTGVAIDPSQITKTEWSAKYFENGNTILHSGY 1083
DB 1050 TDFQKYGGAFDLKAKYPEIFERVQISNGQKMTTEKIYKSAKYFNGTILHSGY 1109
QY 1084 VLKADGGQYNYLGTGTFKQFLPQLTGKQGNQGNQYFYDLAGMVKVNTFIED 1143
DB 1110 VLKQWASNDYLTNRNGEVLVLPQLV--NKNSTYGEVSDANGT-KFYSTSGVQAKNSIQD 1166
QY 1144 SVGNWYFDDGKAVENKHFVDVDSYGEKGY-FFLKNGVSFRGLVQTDNGT-YYPDNY 1201
DB 1167 ENGNWYFDDKRSYLVLTGAHEID-----GKHVYFLKNGIQLRDSIREDENGNQYYVDQ 1219
QY 1202 GKVRNOTINA-----GAMIVTL-----DENGKLIK 1227
DB 1220 GAQVLYNRYTTDGGQNNRYFDAQGYAARGLVKIGDGQOFFDENGQVQK 1266

RESULT 15
Q55263
ID Q55263 PRELIMINARY; PRT; 1590 AA.
AC Q55263
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE GTF-1.
GN GLUCOSYLTRANSFERASE.
OS Streptococcus sobrinus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1310;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33478;
RA Sato S.;
RT "DNA and amino-acid sequences of water-insoluble-glucan synthetase
RT produced from Streptococcus sobrinus ATCC 33478.";
RL Ann. Kagoshima Univ. Dental School 16:23-29(1996).
DR EMBL; D63570; BAA0592.1; -.

```

[illegible]

Db	823	DSRVNFGFSGNFOSFATKEEEXTNVVIANNVNDKVFYSWGIOTDFEMAPQTVSVSTGQFLDSV	882
QY	936	IDNGYAFTRDYDLGMSPEPKYGTDELRNAIQALHAKAGLOVMADWVPDQIYNLPGKEWAT	995
Db	883	IQNGYAFTRDYDLGMSKANKYGTADQLYKAIKALHAKGLKYMADWVPDQMTFFRQEVVT	942
QY	996	VTRVDDRGNVWKDAIINNLLYVNT-IGGGEYKKYKGAFDLKLIKLYPEIFTKKQVSTG	1054
Db	943	VTRTFDFGKPIAGSQINHSLYVTDTKSSGDDYQAKYGAFLDELKEKYPEEFTKKQISTG	1002
QY	1055	VAIDPSOKITPEMSAKYFNCTNIIHRSGSYVLKAD-GGQYNYNLGTTTKQFLQLTGKKQ	1113
Db	1003	QAIDPSVKIIQWSAKFYENSGNIIHGRGADYVLDQSVNKFNFVADT-LFLPSSLGKYVE	1061
QY	1114	GNRGFKVNDGNYFYD--LAGNMVAKTFIEDSVGNWYFFDQDGRKVNKKHFVDVDSYGE	1171
Db	1062	SGIRY----DGKGYIYNSSAVGQVKASPIEA-GLNYFFGKDGVMVYGAQTIN----	1110
QY	1172	KGTYFFELKNVSPRG-----GLVQ	1190
Db	1111	GANYFFLENGTALRWYTYDAGNSHYANDSKRVENGVOQFGNDWRYFKDGNMVAGLTT	1170
QY	1191	TONGTYFYDNYGKVRNQRIT--NAGAMIYTLDENGKLKASNSD	1233
Db	1171	VDGNVQYFEDKGVQANDKIIVTFDGRKRVYFDQHGNGVNTFIAD	1215

RESULT 16

Q59983 PRELIMINARY: PRT; 1590 AA.

AC Q59983;

DT 01-NOV-1996 (TrEMBLrel. 01, Created)

DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)

DY 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)

DE GLUCOSYLTRANSFERASE-1 PRECURSOR (EC 2.4.1.5).

GN GFI.

OS Streptococcus sobrinus.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;

OX Streptococcus.

OC NCBI_TaxID=1310;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CM2176;

RA MEDLINE=94146405; Pubmed=8312602;

FX Sato S., Inoue M., Handa N., Aizawa Y., Isobe Y., Katayama T.;

KT "DNA sequence of the glucosyltransferase gene of serotype d

KT xt Streptococcus sobrinus";

RL DNA Seq. 4.19-27(1983);

RE ENBL: D13858; BAA02976.1; -

DR InterPro: IPR002479; CW.Binding.

DR InterPro: IPR003318; Glyco_hydro_70.

DR Pfam: PF01473; CW.Binding_1; 16.

DR DRAM: PF02324; Glyco_hydro_70; 1.

KW Signal; Transferase; Glycosyltransferase.

FT SIGNAL 1 38 POTENTIAL.

FT CHAIN 39 1590 GLUCOSYLTRANSFERASE-I.

SQ SEQUENCE 1590 AA; 175955 MW; C3C83A57CF3C2B0E CRC64.

```

Query Match      34.4%  Score 2284;  DB 2;  Length 1590;
Best Local Similarity 42.3%  Pred. No. 6.8e-97;
Matches 522;  Conservative 181;  Mismatches 384;  Indels 148;  Gaps 36;

QY      10  AAVYPSLKNASVSGEDFTTKLTNQYQALNGQLVLLRFSKAADGNPSGNTVTDQ---- 65
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      24  ATMLASALGASVASADTD-TASDSNQAV-----VTGQTTNNQATDQTSTA 69

QY      66  ---FSKNVATTGGNFYVYKVRNGNVYFSGHWATNQSNDKDSQMIILVLRGKVKRQLVND 122
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      70  AVATSSQSASTDAATDQASA-AEQTQ-----GTTASTDTAAQ-----TTTNAEKW----- 115

QY      123  TREGAAGFNRNDVYKVNPAIENSSMSFGIITLPTVTKNENQVLVHRFSNPKYKTCGNY 182
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

```

Db 116 -----VPTENEN-----QGFTEMLAEAKNV 136
QY 183 VDFWSELPVDFKFGNGKQKFGLOQTINGQYIDPTTGQPRKNFLIQSNWYIFDS 242
Db 137 ATAESDSIP-----SDLAKMSNVKQVDGKYIID-QDGNVKNFAVSGDKIYYD- 186
QY 243 DTGV--GTALELQFAKTVSSNEQ-YRNGNAAYSDYDKSENWNGYLTADTWYRPKQIL 299
Db 187 ETGAYKDSKVDADRSSAVSQNATIFAANFRAYSTAKNFENYLTADTWYRPKSIL 246
QY 300 KGTWTDTSKEDMPILMWMPNTLTQAYLYNWKHGNLLPSALPFENADADPAELNH 359
Db 247 KDKWTESGKDDFRLMAWPDDETKNYVYM----NKVVGDKTYTETSADLTA 302
QY 360 YGEIQQNIETKRISETGNTDMLTLMHDFVTNNPMWKNDSNVNFSGIFQGGFLKYEN- 418
Db 303 AELVQARIEQKITSENNTKMLREALSAFVKTPQWNGESEKPYD--HLQNGALLFQDQ 360
QY 419 SOLTYPANSYDLRLGMPINIKDQ-----TYR-----GOPELLANDINSNPVQAOL 467
Db 361 TDLTPTQSYNRLARTPTNQTGSLDSRFTYNNPNDPLGGYDFLLANDVNSNPVQAOL 420
QY 468 NLYXLYLNGFTTANNDQANFSDVVDADPNDTADLMNTAODYFNAAYGMD--SDAVSNKH 526
Db 421 NMLHLLNFGSIYANDADANFDSIRVDVNDVADLQLSSDILKAAAYGIDKNNKANNH 480
QY 527 INILEWHDADPEYFNKIGNPOLTMDDTIKNSLNHGLSDATN-RWGLDAIVHQSADREN 585
Db 481 VSIWEASNDPTYLHDDGDLNMDNKNFRLSLMWSLAKPLDKRSLNPLIHNLSVDREV 540
QY 586 NSTENVIPNYSFVRAHDNNSQDQIONAIR-DVTGKY-HITTFEBQKGDIDAYTQDQNS 643
Db 541 DOREVETPVSIFARAHDSQVDIIRDKAEINPNSFGYSFTQEEIQAFKIYNEDLKK 600
QY 644 TVKYNLYNIPASAYALLNKTIPRVYGDLYTDGGYMEHOTRYDITLNLKSRVKY 703
Db 601 TDKKTHYNPLSYLLNKGSIPIRVYGDFTDGGYMAKNTVYDAIESLLKARAKY 660
QY 704 VAGGSMQPMVSGGNNILTSVRYGKGAMTADTGETRQTGIGVYVSNTPNLKLVND 763
Db 661 VSGGQAMQYQI-NGEILTYSRYGKALKQSDGATDRTSGVGMVGMQNFSL--DG 717
QY 764 KYV-LHGAHKNQYRAVLTGTGVIWNTSDQAP-----VAMTDENDLYLSSHNLV 818
Db 718 KYVALMGAAHQAQETRALWYSTKDGATVATADAKAKAGLVKRTDENGILYFLNDL-- 775
QY 819 NGKEADTAVQGYANPDVSGYLAVVVPVPGASNDQARTAPSTEKNSGSAYRTNAFDSN 878
Db 776 -----KGVANPQVSGFLQVVPVGAADQDILRVAASDTASTDGKSLHQDAAMDSR 825
QY 879 VIFAFNPNVYPTKESERANVRIQADFFASLGTSTFEMAPQYNSKDRFTLSTIDN 938
Db 826 VMFEGSNFQSFATKEETNTVIANVNDKPYSWGITDFEMAFQYVSSYDGGFLDSVION 885
QY 939 GVAFTDRYDLGHSSEPNKYGIDEDRNAIQALHAKAGLVWADVPQIYNLPGKEVATVR 998
Db 886 GVAFTDRYDLGHSKANKYGTADLVKAIKALHAKGLKLVWADVPQIYNLPGKEVATVR 945
QY 999 VDDRGHVWVDALINNLYVYNT-IGGGEYQKRYGGAFBLDKLQKLYPEIFTRKQVSTGVAI 1057
Db 946 TDKFGRPIAGSQINHSLYVTDTKSSGDDYQAYGGAFBLDEKXYELTFKQISTGQAI 1005
QY 1058 DPSQKTESAKYFNGTNIHRSGVVLKAD-GQYNNLTGTTKQLPQLTGERKKQGNH 1116
Db 1006 DPSVKIKOMSAKYFNGSNTLGRGADYVLDQVSNKYNFVASDT-LFLPSLLGKVVESGI 1064
QY 1117 GFVKGNQNYFYD--LAGNMYKNTFIEDSVGNWYFFDQDGKVENKHFVDVDSYGEKGT 1174
Db 1065 RY-----DGKGIYINSATGDQVKNASITEA-GNLYYFGKDGAVTGAQTIN-----GAN 1113
QY 1175 YFFLKNGVSRGGLVQTD--NGTYFDNYTGKVMRN 1207
Db 1114 YFFLENGTALR-NTIYTDAGNSHYANDGKRYEN 1147

RESULT 17
Q48756 PRELIMINARY; PRT; 1290 AA.
ID 048756
AC 048756;
DT 01-NOV-1996 (TREMUREL. 01, Created)
DT 01-NOV-1996 (TREMUREL. 01, Last sequence update)
DT 01-JUN-2001 (TREMUREL. 17, Last annotation update)
DE DEXTRANSURASE.
OS Leuconostoc mesenteroides.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
OC Leuconostoc.
OX NCBI_TaxID=1245;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NREL B1299.
RX MEDLINE-97136686; PubMed-8982063;
RA Monchois V., Willemot R.M., Remaud-Simeon M., Croux C., Monsan P.;
RT "Cloning and sequencing of a gene coding for a novel dextranucrase
from Leuconostoc mesenteroides NREL B-1299 synthesizing only alpha (1-
6) and alpha (1-3) linkages.";
RL Gene 182:23-32(1996).
DR EMBL; U38181; ABA0875.1; -;
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 11.
DR Pfam; PF02324; Glyco_hydro_70; 1.
SQ SEQUENCE 1290 AA; 145590 MW; 3555C2E96B749FAA CRC64;

Query Match 34.0%; Score 2256; DB 2; Length 1290;
Best Local Similarity 46.5%; Pred. No. 1e-95;
Matches 471; Conservative 155; Mismatches 335; Indels 52; Gaps 20;

QY 211 INGQYQYIDPTGQPRKNFLIQSGNNWYIFDSDTGVTGTALELQFAKTVSSNEQYRNGN 270
Db 7 VDGKVIYFG-DGQPKKNTTIDCKPYFDKDTGALSNN-DKQYVSELSFIGNKH--N 61
QY 271 AAYSDKRS IEENVNGYLATDYRPRKQILKDGTTWTDKETDMPRLMVMWPNLTQAYY 330
Db 62 AVYNTSSDNFTQLECHLTASSWYRPRKQILKNGKRWAPSTVYDFRELLMAWPPDKSTQVY 121
QY 331 LNYMKQGNLPSALPFENADADPAELNHYSEIVQONTEKRISETGNTDMLTLMHDFVT 390
Db 122 LNYMKDQG--LLSGTHFSDENHETLTAAAMQAQVNTIEKKIGQLGNTDMLKTAQYID 179
QY 391 NPMWKNDSNVNFSGIFQGGFLKYENSDLTYPANSYRLLGRMPINIKDQ---TYR-- 445
Db 180 ACPNNDISEAKGDD--HLOGGALLYTHSDMSFPAANSYRLLSRTPKNGKGIADKYQG 237
QY 446 GQEFFLLANDIDNSNPVQAOLNLYLLNFTGITANNDQANFDSVRVDAPNDIDADLMN 505
Db 238 GPELLANDVONSNPVQAOLNLYLWNNIGSLQNDQANFQDQYRYDAVDNDADLLQ 297
QY 506 IAQDYFNAAYGMDSDAVSNKHINILEDNHADEP---YFNKIGNPQLMDDTIKNSLHNG 562
Db 298 IAGEAKAAAYGVX--MTRERINITYQFWKTGEMKIQTSKHNATSKLSMDPFLHLAIKYA 355
QY 563 LSDATN-RWGLDAIVHQSADRENSTENVIPNYSFVRAHDNNSQDQIONAIRD----- 616
Db 356 LNPNDKRSGLTEPRHSLVKRIYTDKENVAQPNYSFIRAHDSQVQTIADIKDKINPA 415
QY 617 VTGKDYHTTFEDEQKGDIDAYTQDQNSVTYKYNLYNIPASAYLLTNKDTIPRVYGDLY 676
Db 416 STGLD-STVTLQDIQAFEDIYNADELKADKVYTPYTNIPASAYLLTNKDTIPRVYGDLY 474
QY 677 TGGQYMEHOTRYDITLNLKSRVKYVAGGSMQMTSVGGNNILTSVRYGKGAMTADT 736
Db 475 TDDQYMAKQSPYQAIIDALLKARIKAYAGGQTKMKNYFPDQSQVTSYRYGKGAMTAD 534
QY 737 TGTDETRTQIGVWYVSNTPNLKLVNDKVLVLMGAAHKNQYRAAVLTGQVINYTSDQ 796
Db 737 TGTDETRTQIGVWYVSNTPNLKLVNDKVLVLMGAAHKNQYRAAVLTGQVINYTSDQ 796


```

Db 535 SGNQETRYQIGIGLVNRRPDLKLSKDEKMDGAAHKNQDYPVLLTTKSGLKVYSTDA 594
QY 797 GAPVAMTDNDGLDYLSSHNLVYNGKEEADATVQGYANPDVSGYLAVWVPVGSADMDART 856
Db 595 NAFVVRTDANGOLTF-----KAD-MVYGVNDPQVSGYIAAVPVGASEMDART 642
QY 857 APSTENSGNSAYRTNAAFDSNVIFEAFFNFYTPPKESERANVRIAQADYFASLGFTS 916
Db 643 KSETTSTGVSVHSNALDSQVYEGFNFQDPPTTDEFTNKIAQNVNLFKONGITS 702
QY 917 FEMAPQNSKDRFTLSDIDNGAFTDRDYDLGMSPNKYGTGDEDLNAIQALHRAGLAV 976
Db 703 FEMAPQYRASSDKSFLDALVQNGAYFTDYDGYNTPTKYGTADNLDAI RALHGGOGIA 762
QY 977 MADWVQDYLNPCKEAVATVTRVDDRGVWKAIDALINNNLYVNTTIGGGYQKKYGAFID 1036
Db 763 INDWVDPQIYNLPDEQLVTAIRTDGSDHTYSGSDHTLYASKTVAGTYQOQYGGAFLE 822
QY 1037 KLOKJLPEITFKQVGTGVALDPQSKITWSAKYFNGTILHRGSGYVLKADGGQYINLG 1096
Db 823 OLKTQYLPQLFQKQISTDQPMNPDQIKSWEAKYFNGSNLQGRGAWYVLKMGWTOQYFV 882
QY 1097 TTTKQFLPQLAGEKKQGNFV-KGNDGNYFYDLAGNVMVMTFIEDSGNWFYFDDQG 1155
Db 883 SDAQTELPQLLGEK--AKTGFTVTRGKETS--FYSTSGYQAKSAFICDN-GNWFYFDDKG 937
QY 1156 KMYENKHFDVDSYGEKG-TYFPLKNGVSFRGLVQTDNGTYFYFDNYGRMVAN 1207
Db 938 KMYVGNQVIN-----GINYFLPGLIELQDALYLDHGMGYIYNNIGKQLHN 983

RESULT 18
Q9WXJ4 PRELIMINARY; PRT; 1338 AA.
ID Q9WXJ4
AC Q9WXJ4
DT 01-NOV-1999 (Tremblrel. 12, Created)
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE GTF-S.
GN GTF-S.
OS Streptococcus ericeti.
OG Plasmid pAMI.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1333;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HS-6;
RA Inoue M., Fukui K., Miyagi A.;
RT "S.cricetus glucosyltransferase(gtfs and gtf) genes.";
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB026123; BAA77236.1; -.
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_3; 10.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Plasmid.
SQ SEQUENCE 1338 AA; 148558 MW; 0A90C8E10E15D99B CRC64;

Query Match 32.8%; Score 2180; DB 2; Length 1338;
Best Local Similarity 43.4%; Pred. No. 3.3e-92;
Matches 483; Conservative 154; Mismatches 370; Indels 106; Gaps 23;

QY 208 LQTINGQQYIDPTTQCPKRNFLQSGNNWIFDSDFGVG-----TNALELQFAKGYTS 261
Db 120 LENWDGKTYVD-ANGQRUKNSTVTDGTYFYFDAGTQGAQAEPTQINQNDNQVADPTYA 178

QY 262 SNEQYRNGNAAYSDDKSTENAVGYLTADTWYRPKQILKDGTTWTDSKETDMRPIILWVW 321
Db 179 AN-----NQAFNDVSSRETQVNYTASWPRPKILKNGESQWASAESDMRPIILWVW 232
QY 322 PNTLQAYLYNMKQGNLLPSALPFNADADPAELNHYSEIVQOQNIKRISETGNTDWL 381

```

RESULT 19

Q55265

ID Q55265

AC Q55265;

DT 01-NOV-1996 (Tremblrel. 01, Created)

PRELIMINARY; PRT; 1577 AA.

RESULT 19

Q55265

ID Q55265

AC Q55265;

DT 01-NOV-1996 (Tremblrel. 01, Created)

PRELIMINARY; PRT; 1577 AA.

RESULT 19

Q55265

ID Q55265

AC Q55265;

DT 01-NOV-1996 (Tremblrel. 01, Created)

PRELIMINARY; PRT; 1577 AA.


```
DT 01-NOV-1996 (TREMREL. 01, Last sequence update)
DT 01-JUN-2001 (TREMREL. 17, Last annotation update)
DE GLUCOSYLTRANSFERASE PRECURSOR.
GN GRM.
OS Streptococcus salivarius.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1304;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95122197; PubMed=7822030;
RA Simpson C.L., Giffard P.M., Jacques N.A.;
RT "Streptococcus salivarius A9C 25975 possesses at least two genes
RL coding for primer-independent glucosyltransferases.";
DR Infect. Immun. 63:609-621(1995).
DR EMBL; L35928; AAC41413.1; -
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 11.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Signal; Transferase.
FT SIGNAL 1 38 POTENTIAL.
FT CHAIN 39 1577 GLUCOSYLTRANSFERASE.
SQ SEQUENCE 1577 AA; 175290 MW; 3EF898A7D3A7BF3 CRC64;

Query Match 32.8%; Score 2175.5; DB 2; Length 1577;
Best Local Similarity 40.4%; Pred. No. 6.5e-92;
Matches 518; Conservative 189; Mismatches 415; Indels 159; Gaps 38;

QY 19 SAVSGFDTHIKLTN---DQYQALNGQLVLLRFSAADGNPSGDNF-VTDQPSKNVAITG 74
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 75 GNFDYKVNQNGQVFCGWHATQSNQDKSCWIIIVN-GKEYKRLQVNDTKGGAAGFN 133
DB :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB 146 -----QVAGQTSQAQTPSVTEQARP-----VLTNAAPAATRAADSTIRINA--NRN 191
QY 134 DVYKVPALNENSMGFGIITLPVTVK-----NENVLVHREFSNDVKTEGNY 182
DB ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB 192 ----TWITASGTPNVIITGPNTPKPNVVTSPNGRPNVTIVTQNPQN----- 241
QY 183 VDFWSELMPVKDSF-QKNGPLK-----QFGLQFINGQQYIIDTPGQPRKNFL 230
DB ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB 242 -----PVQPSQSPQNPQPNQPSLDYKPVASNLKTIQKQYVE--NGVYKKNAA 291
QY 231 LQSGNNWIYFDSGTGTHALELPKAGTVSSNEQYRNCNAAAYSDDKSIENVNGVLPAD 290
DB :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB 292 IELDGLXYFD-ETGAMVDQSKPLYPADAIPNNSIYAVYNQAYDTSKSFELHDLNFLTAD 350
QY 291 TWYRPOILKDGTTTDSKETDMRPIKLVWNPNTLTQAYLVNMYKOHGNLPSALPFNA 350
DB :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB 351 SWYRPOILKDGKNNTASTEKDYRPLLTWPDKVTQVNYLYMSQQG----FGNKTYT 406
QY 351 DADPAELNHYSEIVQNIETKRSCTGNTDWTMLTLMHDFVTNNPMMNKDSE--NVNFSGIQF 409
DB :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB 407 DMHSDYLAATAETVQIBERIGRENTWLRQLMSDFIKTOPGWNSEEDNLLVKGDL 466
QY 410 QGFLKYNSDITPVANSYRLLGRPNIK-----DQYRGQEFLLANDIDNSNPV 462
DB ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB 467 QGGAFLNNSATSHANSDFRLMNTPTNQTTRKYHIDRSNGGYELLANDIDNSNFAV 526
QY 463 QAEQLNWLXLLNFGTITANNQANFSDVSDVPADPNIDADLNIAQDYVNAAYGM--DSDA 521
DB ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB 527 QAEQLNWLHYINNIGSILGNDSANFDGVIDAVDNVDADLIQASDFEKEYRVADNEA 586
QY 522 VSKNHNILEDNNHADPEYFNKIGNPQLTMDPTIKNS--LNEGLSDATNRWGLDAIVHOSL 580
DB :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB 587 NAYALHLSLEANSYNDHQVKNKDTKGAQLSNDPQRETLLTTLRKSNYRGSLEVTNLSL 646
QY 581 ADRENSTENVIPNYSFVRHNDNSQDQIQNAL-RDVTGK--DYHTTFEEDQKIDAYI 638
DB :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB 647 NNRSEKQHTPRDANTIFVRAHDSVQAVLANIISKQINPKPTDGTFTMDLQKAFETYN 706
```

QY	1047	TKQVSTGVAIDPSQKITEWSKAYPNFNGNINLHRSGYVYLKADG-QQYVNLGTTTKQKFLPI	1105
DB	1023	TAKMISTGHPIDSSSVKLKWSAQYFQVNLGRDYVYLSDEGTGKYFTV-NEKGBFLPA	1081
QY	1106	QLTGEKKQGNIEGFKVNGDGNYYFYDLAGMVKNTFIEDSVGNWYFFDODGKMVENKHPVD	1165
DB	1082	VLTGD-KEATGDFYNDGKWTFF-TTASQAKSDFV-TVAGNTYIFDYIGHVWYGPNGIN	1138
QY	1166	VDSYGEKGYFFFLKNGSVPRGLVOTDNG--TYY	1197
DB	1139	TK-----FYFLPNCVLMKDAVMDDDRSGSVYKGTGVMYKSGRNNWFAMTDSKQLR	1192
QY	1198	---FDNYGKM-VRNQTINGAMYYITDENGKLIKAYSNDREYPTSTDVGKMLDQN	1249
DB	1193	FRHFDNYGMSVLVTHGNVOYY--DEGFOVKGDFVTDKAGOT-----RYFDKN	1241
RESULT	21		
Q00600			
ID	Q00600	PRELIMINARY; PRT; 1518 AA.	
AC	001600;		
DT	01-NOV-1996 (TrEMBLrel. 01, Created)		
DT	01-NOV-1996 (TrEMBLrel. 01, Last sequence update)		
DT	01-JUN-2001 (TrEMBLrel. 17, last annotation update)		
DE	GLUCOSYLTRANSFERASE I (EC 2.4.1.5) (GTF) (DEXTRANSUCRASE).		
GN	(SUCROSE 6-GLUCOSYLTRANSFERASE).		
GN	GTFJ.		
OS	Streptococcus salivarius.		
OS	Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;		
OX	Streptococcus.		
OX	NCBI_TaxID=1304;		
RN	[1]		
RC	SEQUENCE FROM N.A.		
RC	STRAIN=ATCC 25975;		
RC	MEDLINE=92148377; PubMed=1838391;		
RA	Giffard P.M., Simpson C.L., Milward C.P., Jacques N.A.;		
RA	"Molecular characterization of a cluster of at least two		
RT	glucosyltransferase genes in Streptococcus salivarius ATCC 25975.";		
RL	J. Gen. Microbiol. 137:2577-2593(1991).		
CC	-1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT TO		
CC	PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE OF		
CC	THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE		
CC	AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.		
CC	-1- CATALYTIC ACTIVITY: SUCROSE + ((1,6)-ALPHA-D-GLUCOSYL)(N+1).		
CC	PRUCTOSE + ((1,6)-ALPHA-D-GLUCOSYL)(N+1).		
CC	-1- SUBCELLULAR LOCATION: EXTRACELLULAR.		
CC	-1- DISEASE: DENTAL CARIES.		
DR	EMBL; Z11873; CAA77900.1; -.		
DR	EMBL; M64111; AAA26896.1; -.		
DR	InterPro; IPR002479; CW_binding.		
DR	InterPro; IPR003318; Glyco_hydro_70.		
DR	Pfam; PF01473; CW_binding_1; 13.		
DR	Pfam; PF02324; Glyco_hydro_70; 1.		
KW	Transferase; Glycosyltransferase; Repeat; Dental caries.		
FT	DOMAIN 1307 1482 6 DIRECT REPEATS.		
FT	REPEAT 1307 1338 REPEAT 1.		
FT	REPEAT 1339 1352 REPEAT 2.		
FT	REPEAT 1372 1403 REPEAT 3.		
FT	REPEAT 1404 1417 REPEAT 4.		
FT	REPEAT 1437 1468 REPEAT 5.		
FT	REPEAT 1469 1482 REPEAT 6.		
SQ	SEQUENCE 1518 AA; 167730 MW; DAA41F710789859A CRC64;		

```
Query Match      31.5%; Score 2093; DB 2; Length 1518;
Best Local Similarity 37.2%; Pred.No. 3.8e-48;
Matches 511; Conservative 197; Mismatches 414; Indels 250; Gaps 42;

Qy    27   T K L I N D O Y Q A L G Q L V L R F S K A A D N P G D M T V T D Q F S K N Y A T T G G N F D Y K V K N G N Q 86
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    18   T I A V A S V A L T V I G G L S V T --- T S S V S A D E T Q D K T V T C O S N S G T T A S I V I S P E A T K ----- 69
```

843	SYTASALDULIYEGFSNFUDFQKUAQYTNKIAENTDLPKAGVTSFEMAPQVYSAY	902
928	DTFLOSTIDNGYATFTRDYDGLGMEPNKYGTBDLNRNQAQLHKAQLOWADNPDQTYN	987
929	: : : : : : : : :	
903	DCYFDSLITONGYAFSDRYDLAMSKNNKYSGKEDLANALKAHAGIAIADNPDQYIQ	962
988	LPCKEYATVTRVDRNGWKDAIINNNLYVYNTIGGS-EYQKYYGAGFLDKIQKLYPEIF	1046
963	LPKEVYVTSADVNTGRVKIDOPMWNKLYANTKSSGKGFQAKYGEFTLAELQYQPEMF	1022

Matches 449: Conservative 157; Mismatches 370; Indels 79; Gaps 25;

```
QY 212 NCQOYYIDPTTGQRKNEFLQSGNNWYFDSDTGVGRNALELQFAKGTVSSNEQYRNGNA 271
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 175 DGKYYL-LEDGSHKKNFAITVNGQVLYFD-ENGALSSTSYSTQETNLTVDFTKNA 232

QY 272 AYSYDDKSIENGVNGLRADTWYRKQILKDDTTWDSKETDMRPILMVWNPNTLTQAYYL 331
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 233 AYDSTKASFELVDYLTADSWYRKELIAGTTWKASTKDEKFRPLASWNPDKUTQAYL 292

QY 332 NYMKQHGMLPSALPFFNADADPAELNHYSEIVQONIEKRISSETGNTDMLKTLMHDFVTN 391
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 293 NYMKALNSGETKDVFTIENSQSLSNAAAILQKKEVKTAAKSTDMLRQSTAEAFVKD 352

QY 392 NPMNKKDSYNEFSGIOFGFLPYNSDLPYANSYRLLRMPIN-----IKDQTYR 445
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 353 QDKWNINSEPGKE--HFQKALLFVNSDSYKWNANSYRKLNQATSYIKNKHIVNGSDG 410

QY 446 GQEFLLANDIISNPVQAEOQLNLYILLNFETIT--ANNQOANFDSVYRDAPDNIDADL 503
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 411 GYEFLLSNDIISNPVQAEOQLNLYIPMNQGVFGDKDAHFDGIRVDADVNSVDM 470

QY 504 MNIADYFNAAYGM-DSDAVSNKHINILEDNHADPEYFNKIGNPQLTMDTINKSNLHG 562
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 471 LQLVSSYKAAKYNESEARALANISILEASHNDPYVYNEHTAALSMONGRLSIVHG 530

QY 563 LSDARNRGLDAIVHQSLAD-----RNNSTENVVPIPNYSPVRAHDNNSQDI 610
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 531 LTRPVNKGTCGA-RNASMKDLINGGYFGLSNRAEVTSYDQLGFATYLFVRAHDS-----EV 585

QY 611 QNAIRDVTK-----DYHTTFEEOEKGIDAYIQDQNSTVKKYNLYNPASAYAILTNK 664
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 586 QTVIADIISKIDPTDGTFTLQKQAFDIYNADMLKVDKETHSNIPAYALMLQTM 645

QY 665 DTIPRVYCDLTGQGMERQTRYDFTLNLKSRVYKAGGSMQTSYGG-----717
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 646 GAATRVYCDLTGQGMERQTRYDFTLNLKSRVYKAGGSMQTSYGG-----717

QY 718 -NNNLLTSVRCKGKATADTR-GTDETRTQIGVVVSTPNLKLGVNDKVVHLHGAHKN 775
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 706 DNKEVLVSVRYGQDLMKSTDEGGKYGRNSGMILTANNPDLKLADGETITVNGAHLN 765

QY 776 QYRAAVLTITDGTGVTINTSDQAP--VAMTMDENGDLVLSHNLVYNGKEADTAQGVANP 834
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 766 QAYRPLLLGTEKGVSSLNDSDTKLVKXTDAQG-----NLVTTADE-----TKGFTV 813

QY 835 DVSGYLVAVVVPVPGASDNDARTAPSTEK-NSGNSAYRTNAAFDNSVIFEAESFNFTPTK 893
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 814 DMSGYLSVWVPVPGATDQNVLAKPSTKAYKEGDKYSSSALEAQVIEGFSNFQDFVKE 873

QY 894 ESERANVRIANADFFASIGFTSEMAPOYNSKDKORTFLDSTIDNGVAFTRDYDLGMS 953
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 874 DSQYTNKLIAANADLFKSWGITSFEIAPQYVSSKDGTFDLSIIENGAVFTDRYDFAMSKN 933

QY 934 NKYGDEDLNRAIQALHRAQLQVADVPDQIYNLPKGEVATVTRVDRGNVWMDALINN 1013
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 934 NKYGSKEDLRALALHKGQIQVIADWVPDQIYLPKGEVATVTRVDRGHVLDLDTSLVN 993

QY 1014 NLYVNT-IGGGEYQKYGGAFLDKLQKLYPELFTKQVSTGVAIDPSQKTTESAKYFN 1072
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 994 KLYVTNTRKSSGNDFAQYCGGAFLOKLQKLYPEIEKEVMEASGKTIDPSVKIKQWEAKYFN 1053

QY 1073 GPNILHRGSGVYLKADGGQYXNL---GTTTQFLPIQLTGEKKQCGNEGVAGNDCNYYFY 1129
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1054 GTNIOKRGSDYVL-SDGKLYFTVNDKGF-----FLPAALTGDTK-AKTGFAYDGGVTVY 1106

QY 1130 DLAGNNVANTFEDSVGNWYFFDDGKRVENKHFVDVDSYGEKGTFFPKKNGVSFRGLV 1189
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1107 TTSCTQAKSQFTYIN-GKQYFENDKGYLVTGEQITD-----GSNYFFLPNGVMTDQVR 1159

QY 1190 QFDNCTGYFDNKGMRVNOTINAGAMHYTLDENGK 1224
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1160 KNAKGQSLV--YKSGKLTQTQGMKEVTVKDDSGK 1192
```

```
RESULT 23
Q91CJ7 ID Q91CJ7 PRELIMINARY; PRT; 1016 AA.
AC Q91CJ7;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE DEXTRANSUCRASE.
GN DSRT.
OS Leuconostoc mesenteroides.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
OC Leuconostoc.
OX NCBI_TaxID=1245;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL B-512F.
RX MEDLINE=20169623; PubMed=10705445;
RA Funane K., Mizuno K., Takahara H., Kobayashi M.:
RT "Gene encoding a dextranucrase-like protein in Leuconostoc
RL mesenteroides NRRL B-512F.";
RL Biosci. Biotechnol. Biochem. 64:29-38(2000).
DR EMBL; AB020020; BAA90527.1; -.
DR HSSP; P06278; 1VJS.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF02324; Glyco_hydro_70; 1.
SQ SEQUENCE 1016 AA; 110344 MW; 8896BFDE13CCB47 CRC64;
```

Query Match 30.0%; Score 1989; DB 2; Length 1016;
Best Local Similarity 50.7%; Pred. No. 1.4e-83;
Matches 407; Conservative 100; Mismatches 252; Indels 44; Gaps 10;

```
QY 208 LQTINGQOYIDPTTGQRKNEFLQSGNNWYFDSDTGVGRNALELQFAKGTVSSNEQYR 267
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 230 LKTDIGKTYIYD-DDQVKNFATVIGKLYFDKEIGALADNDYQFLEGLISENTYTT 288

QY 268 NGNAAYSDDKSIENGVNGLTADTWYRKQILKDDTTWDSKETDMRPILMVWNPNTLTQ 327
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 289 EHNASVGTTSYSDTYNVDGYLTADSWYRKQILVNGQWSSKDDDLRPLLTWWPDKATQ 348

QY 328 AYLYNKHQHGMLPSALPFFNADADPAELNHYSEIVQONIEKRISSETGNTDMLKTLMH 387
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 349 VNYLNAMKYLDAETETEV--YTSDDSQDALNKAQNTQVKIEKISQEGQTQNLKDDISK 406

QY 388 FVTNPMWKNKDSNVNFSGIQFGGFLKYENSOLTPYANSYRLLRMPIN-----IK 440
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 407 FVDSQSNWNIASESQGTG--HLQGGALLYVNSDKTPANSYRLLNRTPTNQTGTPLYTT 464

QY 441 DQYTRGQEFLLANDIISNPVQAEOQLNLYILLNFQGITANNDQANFDSVYRDAPDNID 500
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 465 DPTQGGYDFELLANDVDNSPNVQAEOQLNMYILLNFQGITNNDADANFDSIRYDAVDND 524

QY 501 ADLNIADQYFNAAYGMD-SDAYSNKHINILEDNHADPEYFNKIGNPQLTMDTINKSL 559
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 525 ADLLQIAADYFKAAYGYDKSDAISNQHVSILEDSNDADAEYVKONGDNQLSNDKRLSL 584

QY 560 NHGLS-----DAFNRWGLDAIVHQSLADRENNSSTENVVPIPNYSEVRAHDNNSQDQ 613
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 585 KYSLTMPAVDQYGNKRSGLPEFLTNSLVDRNTSDNTAQPNYSEVRAHDSEVQTYIAEI 644

QY 614 IRDVTGDKDYHTF--TFDEQKGDIDAYIQDQNSVKKYNLYNPASAYAILTNKDTIPRVY 671
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 645 IKQRIDPDSGLSLPTMQLTAEKRIYNADQLKTDKEPTQYNIPSTVATILTNDKDTVPVY 704

QY 672 YGOLYTPGGQYMEQTRYDFTLNLKSRVYKAGGSM-----QTSVSGNNNLT 723
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 705 YGDMITDGGQYMATKSYIYDAIDFTLLASRIKYVSGGQTSKMYWQGDSSMAADSYRGILT 764

QY 724 SVRYGKGATATDGTDETRTQIGVVVSNTPNLKLGVNDKVVHLHGAHKNQOYRAAVL 783
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 765 SVRYGNGANTATDAGTNETRTQGIATVIESNPNDLKLSSTDQVVDVAGIAHKNAAYRALL 824
```

Qy	784	TTDGVINVTSQQAP--VAMTDGNDGLYLSHMLVWGKEADTAQGYANPDVSGYL	840
		: : : :	
Db	825	TTKDGIDTVYSDVSGSLIRYTSNGQLIFNSSDIV-----GTANPQVSGYL	872
Qy	841	AVWPVGCASQDQARTAPSTEKNSGNSAYETNAFDSNVIPFAFSNFVYTPTKESRANV	900
		: : :	
Db	873	AVWPVGCASQDQARTESSTATVYDQCTUHSNALLDSQVIYSEFNSFSTPTTKAAYANV	932
Qy	901	RIANADFFASLGFTSFEMAPQYTSKDRTEFLDSTIDNGYATFDYDGLCMSEPNKRYGTDE	960
		: : : : : : :	
Db	933	QIANNTDLYKSWGITNFEFPQYRSSTDSSELDSTIQNGAYETDLYDLGENTPKYGTVD	992
Qy	961	DLRNAIQALHKAGIQVWADWVPD	983
		:	
Db	993	QLRTAIRKALHATGIKANADWVPD	1015
RESULT 24			
Qy	784	PRELIMINARY:	PRF: 2817 AA.
AC	Q97K42		
AC	Q97K42		
DT	Q1-OCT-2001 (Tremblrel. 18, Created)		
DT	01-OCT-2001 (Tremblrel. 18, Last sequence update)		
DT	01-DEC-2001 (Tremblrel. 19, Last annotation update)		
DE	UNCHARACTERIZED PROTEIN, RELATED TO ENTEROTOXINS OF OTHER		
DE	CLOSTRIDIALES.		
GN	CAC1079.		
OS	Clostridium acetobutylicum.		
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;		
OC	Clostridium.		
OX	NCBI_TaxID=1480;		
RP	[1]		
RP	SEQUENCE FROM M.A.		
RC	STRAIN-ATCC 824 / DSM 792 / VKM B-1787;		
RX	MEDLINE=21359325; PubMed=11466286;		
RA	Neelling J.; Breton G.; Omeichenko M.V.; Makarova K.S.; Zeng Q.,		
RA	Gibson K., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,		
RA	Tatusev R.L., Sabathe F., Doucette-Stam L., Soucaille P., Daly M.J.,		
RA	Bennett G.N., Koonin E.V., Smith D.R.		
RT	*Genome sequence and comparative analysis of the solvent-producing		
RT	bacterium Clostridium acetobutylicum.*;		
RL	J. Bacteriol. 183:4823-4838(2001).		
DR	EMBL: AE007623; AAK79053.1; -		
DR	InterPro: IPR002479; CW_binding.		
DR	Pfam: PF01473; CW_binding_1; I19.		
KW	Complete proteome.		
SQ	SEQUENCE 2817 AA; 318224 MW; 18510D04FFBE921 CRC64;		

```

Query Match      4.8%; Score 317; DB 16; Length 2817;
Best Local Similarity 20.4%; pred. No. 3.3e-06;
Matches 314; Conservative 185; Mismatches 516; Indels 522; Gaps 86;

QY 13 YPSLYN-----SAVSGEDTTIKLN-----DQYQALNGQLVLLRFSKAADG 54
| | | | | : | | | | | : | | | | |
Db 1142 YSSYFNTYIFNSDGTAKTGETYLNKTYFNPSPDERMLQGYQYING-----HYFAPDG 1196
| | | | | : | | | | | : | | | | |

QY 55 NPSGDNTVTDQFSKNY-----ATTGNGEDVYKNGNQVFESG-----WHATWQSN 99
| | | | | : | | | | | : | | | | |

Db 1197 TWQ-TGWITNGSSKYYLDPSGAATG-----LQTINGKCFSDNGILOHNGIFYGNTYY 1251
| | | | | : | | | | | : | | | | |

QY 100 DKDSQWIIIV-----LVNGKEVKRQLVNDITKEGAAGFNRDVYKVPATENSSMSGFGQIIIT 155
| | | | | : | | | | | : | | | | |

Db 1352 GSDNNGIMLTGLQLLINGLYLCFNSDGSVKTGLVYTLGKTY-----FDSYSVSGFNQI-- 1304
| | | | | : | | | | | : | | | | |

QY 156 LPVTVKKNENQLVHRESND-----VKTGEGNYIVDFWSELPMVKDSFGKMGKPLKQFGLOQTNG 213
| | | | | : | | | | | : | | | | |

Db 1305 -----NNN-----TYIFGNDGTMQTGWNVGY-----RYLNDSGIKVYTGWQITDG 1346
| | | | | : | | | | | : | | | | |

QY 214 QQXYIDPTTGPRKNFLLQSNWNIIYFDSDTGV-----GTWALELQFAKGVSSNEQ 265
| | | | | : | | | | | : | | | | |

Db 1347 NKYYFD-----TYGAKTGVINIDNYGNYGFN-----NSGYVLGTGWQ 1381
| | | | | : | | | | | : | | | | |

```

Qy	266	YRNGAAYSDDKSIENVNGYL	T--ADTW-	-----	RPKQILKDGTYM	-----	305
Db	1382	HING-STYFENSIGIAN	TGCFITLTKTYFDSYGRMIOGSM	TNGTSYFYFANGVMKTS	1439		
Qy	306	TDSKETDRIILAVWPNWT	TQAYILNYMKHQGNLLPSAL	PENADAPAEUNHYSEIVQ	365		
Db	1440	TDSPNT	--LAWG--	--VRUSYYQYILNAAGT	KLTLQTLQIDGNTYFDSN	--GIMQ	
Qy	366	QNIKRISEGN	-----	TDKLRLMHDEVNPNWKNDS	ENYFESQIQOFGFLKYENS	420	
Db	1489	TGI	---ITINGNRYGVGVNML	GLQFINNTYIS	-----	NSYGIS-OTGFVTLGN-	
Qy	421	LTPYANSDYRLGRMPINIK	QDQYRQEBELLAND	IDNSPVYQABLWMLXY	-----	LLN	
Db	1538	-TYFDS-	--YGBMRIGL--	--YIINNYYYP--	-----	NSKGIMETGWISLYAXPNPGLL	
Qy	476	FGTITANNOANFDSYRVD	ADPNIDADLWNIADYFNAA	-YGMDSDAVSNKHINILEDWN	534		
Db	1586	TGFOTINGKTYFENS	-----	DGSLLYDLQ-YINGSYY	FGDKNGVMLYGLQTI	-----	
Qy	535	HADPEYFNKIGNQ	--LTMQ-DTIKNSLNHGL	SATNWGLDAIVHOSLADREN	NSTEN	590	
Db	1632	GGNTYILNSWGISOS	GFILNGKTYFDSYGM	-----	RTGTQNI	-----	
Qy	591	VVIPNWSFYRAHDNN	-----	SQDQIQ	-----	NAIRDVTKGDIHTFTFDEBQK	
Db	1674	---NYFF-	GDNGTLQGTWISODNL	RYVANSSSGVCLTGLQTLQIDGK	KYFENS	YARMETG	
Qy	634	I----	DAYIQDSNFKKYNLYNI	-----	-----	PASVAILLT	
Db	1728	LVIINNTYFGDNDGT	LJLXSWHNINGMYCFNT	DGTVKTGNTINYLGRCYILDSSQ	GFELST	1787	
Qy	663	NKDTIPR	--VYIGDLYT--	-----	DGQYMEHOTRYEDTLNLK	698	
Db	1788	GLLTIGHNIYFSGDS	TSMTGWTYSSSKYFYNESGIMLT	GTQTLQIDGNTYFDSYGN	---	1844	
Qy	699	SRVKYVAGQSM	-----	QTHSVSGNNNILLTSVRY	KGAMTATDGTDE	741	
Db	1845	---STTGTRISNGCY	GFNDGIMLTGMOTIS--	GNNYFNPDCGTAKIGLNTVEG	KTYT	1898	
Qy	742	TRTOG	--IGVYVSNTPNLK	LGVNDKVLHMGAAHKNQOYRA	-----	AVITTTDG	
Db	1899	FSTGGYTQGIINNS	NTYFYGVYDG	--ALKTGWIRNNYIYAD	NNNGIITQTLGLKIDGKNY	1956	
Qy	789	VINYTSQGPVAMT	-----	DENGDLXLSSHNLVYNGKE	-----	EADTAVQG	
Db	1957	LFNVSG	-----	IMMNGIQSINNYY	GFNDGHLGVHS--	INGQTYTFNADGSKYKGTWI	
Qy	831	-----	YANPD-VSGYLA	VWVPVGASDMDARTAF	TEKNSGNSAYRTNAFDSN	WIFE	
Db	2010	PYBCKMYANPSYCT	GFAT-	-----	NNNTYFDDNGAMETG	IITV	
Qy	883	AFS	-----	NVYVTPFKESBRANVRIQ	ONADFFA	-----	
Db	2050	DNSKYINSYGIBET	GKIIDGKTYFENPNYSGIMQ	TVGISLNGNYFYDDNGVWQIGLQ	2109		
Qy	916	SFEMAPQNSKDRTE	FLDSTDNGVAFTDRYDLG	MSG---	PNKYGTDEDLRMI	-----	
Db	2110	KLNGNTYFENS	-----	NGTALQWTLGSKNYFNP	NSPNSGAAGVIES	INGHYT	
Qy	967	-----	QALHKAGLQVADW	NPQTYNLPKGEVATYR	VDGRGNVWKDALINN	LVVN	
Db	2159	YFNNGIMQTVQRING	YV--YCFENDGTAM	-----	NGFQS--	INGNTYILNIYG	
Qy	1020	-----	TICGGYBQKYGCAF	DLKQKLYPIF	-----	TKKQVSTGVAIDPSQK	
Db	2205	CTYTGWQITNS	NRNYFMSDGVAMLTGAQ	NTITGYIYGFDSNGIML	AGVQTLAGNTYDESSNG	2264	
Qy	1063	--ITEM	-----	SAKYFNGTILHRGSGY	LVKADGGQYINL	TTTKQFLPQLTGCKKOGNE	
Db	2265	TATGWILNSKTYFSP	SLGKL/GFI-TVSGD	NYIYLDADGVLTQGTWTV	DGNKYYLNS	2323	

```
QY 1117 -----GFYKGNNGNYFYFDLAGNMVKNFTFIEDSVGNWYFFDQDGKVB-----NKHF 1163
DB 2324 NGVRQTFGLTLNNKYF---DTNGVQGTFTINNWNYYFNDGCMQFLWTIDYNNKY 2380
QY 1164 VD-----VDSYKGGTYFFLKNVSPRGGVLQVDNGTYF-FDNGYKMWNRQTN 1211
DB 2381 FDSGTGRLGFLQID--GKK--YYNSNGIMLTG--VQTIGNKFGFDNGYGYMLTG----2430
QY 1212 AGAMIYTLDENGKLIKASVNSDAEYPTSTVDGKMLDQ 1248
DB 2431 ----VQDIDNH-----TYNFQDGSASTGWTKLSDK 2457

RESULT 25
Q99W47 PRELIMINARY; PRT; 1385 AA.
ID AC Q99W47;
OS Staphylococcus aureus (strain N315), and
OC Staphylococcus aureus (strain Mu50).
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
OX NCBI_Taxid=158879; 158878;
RN [1]
SEQUENCE FROM N.A.
RC SPECIES-S.aureus (strain N315), and S.aureus (strain Mu50);
RA MEDLINE=21311952; PubMed=11418146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,
RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
RA Mizutani-Uji Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,
RA Kanchisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
RT aureus";
RL Lancet 357:1225-1240(2001).
DR EMBL; AF003131; BAB41751.1; -.
DR EMBL; AF003359; BAB56724.1; -.
DR InterPro; IPR001899; Gram_pos_anchor.
DR Pfam; PF00746; Gram_pos_anchor; 1.
KW Complete proteome.
SQ
SEQUENCE 1385 AA; 149643 MW; 2A0CD27773B3C1D CRC64;

Query Match 3.8%; Score 251.5; DB 16; Length 1385;
Best Local Similarity 20.6%; Pred. No. 0.0014;
Matches 240; Conservative 135; Mismatches 381; Indels 407; Gaps 61;

QY 257 KQTVSS--NEQVRNGNAAYSYDDKSIENYNG-----YLTADTWRPKQILKD-GTTWTD 307
DB 93 KEMVSGQNETTSNGK--SIEKESVQSTGTGNKVEVSTAKSDQASPKSTNEDLNTKQPI 150
QY 308 SKETDMRPTLMVWNPNTLTQAYLLNMYKQHGHLLESALPEFFNADA--DPAELNHYSEIQV 365
DB 151 SNQGLQPDLL-----ENKSVVWVQPTNEENKKVDAKTESTTLNWKSDAIK 196
QY 366 QNIEKRISGTNDMLRLTLMHDFVTNPNMWNKDSNVNPFQFGQFGLKYENSQD-LRPY 424
DB 197 SNAE-----TLVDNNS--NSNNEN-----NADIILPK 221
QY 425 ANSDYLLGRMPT-----NIKQTYRGQEFLLANDIDNSNPVQABQLNWLYYL 473
DB 222 STAPKSLNTRMRMAALQPNSTDSKNVND-LTSTNTLTVTVDADNKTIVPAQD-----YL 275
QY 474 LNFQGTITANDQANFDSVRVQAFDNLIDLALMLIAQDYFNAAYGMSDVAENKHMILSDW 533
DB 276 SLKSQITVDKVKSGDYFTIKYSDTVQ-----VYGLNPEDI--KNIGDIKDP 320
```

Search completed: August 12, 2002, 07:59:15
Job time: 316 sec

```
QY 534 NHADEYFNKIGNPOLTMDTITKNSLNHGLSDATNRW-----GLDAIVHQ-----578
DB 321 NNGETIATAK-----HDTANNELITFTDYVDRFNSVKMGINYSIYMDADTIPVDKK 372
QY 579 -----SLADRENNSTENVIPNYSFVRAHNNNSQDQONAIRDVTGKDYHTFTFD-----629
DB 373 DVPEFSVTIGNQITTTADITYPAY---KEADNNS---IGSAFTETVS---HVGNVDEPGY 423
QY 630 -----EQKG-----IDAY-----IQDQNSVTKYKNLYNIPASVAILLTNK--664
DB 424 INQVYVYVNPMDKDLGAKLKYEAHPKYPYTHIGQINQVNIKIYRVEEGITL---NNGY 480
QY 665 --DTIPRYVYGDLYTDGQYMEHQRYVD--TLTN-----LLKSRVYVAGGQ-----SMQT 712
DB 481 DVTNMDLVDVTDDEFKNKMTYGSNOSVNLDFDGIITSAIYVWVNTKFPQYTNSESPTLVQMAT 540
QY 713 MSVGGNNILTSVR-----YKGAMTATDT---GTDETRIQIGG-----VV 750
DB 541 LSSTGNKSVSTGNALGFTNNQSGGAGQEVYKIGNYVWEDTNKNGVQELGKGVGNVTTV 600
QY 751 VSNTPNLKLG--VNDKVVLHMGAAHKNOQYRAAVALTTTDTGVTINTSDQAGAPVAMTDENG 808
DB 601 FDNNTNTAVGEAVTKEDGSLYLPNLPNGDYRVFESNLPKGYEVPKQGNNEEL--DSNG--658
QY 809 LYLSSHNLVYNGKEE-----ADTAVQGVANPDVSGYLAVVWVPGAS 849
DB 659 --LSS-VITVANGKDLNSADLGIYKPKYNLGDIYVWHDNTNKNGTQDQDEKGISGVTV---710
QY 850 DNQDARTAPSTEKNSGNSAYRTNAAFDSNVIFEAFSNVYPTKESERANVRIQNRADFF 909
DB 711 -----TLKDENGWLVKT-----VTTDADGKYKFT---DLDNNGYKV---EFT 746
QY 910 ASLGFTSFEMAPQYNSKDRITFL-----DSTIDNGYAFDTRYDLG---MSEPNKY 956
DB 747 TPEGYTPVTTSGSDIENSLGITTTGVINGADNMTLDSGFYKPKYNLGNYWVEDTNKD 806
QY 957 GTDEDLRNAIQALHAGLOVMAVDPDQIYNLPKGEVATVVRVDRGNVWKDAIINNLY 1016
DB 807 GKQDSTKSGISGV-----TVTLKNNGEV-----LQ 832
QY 1017 VVNTIGGGEYQ--KKYGGAFDLKLQKLYPELTKKQVSTGV--AIDPSQKITENSAXYFN 1072
DB 833 TTKTDKDGKYQFTGLENGTY--KYEFETPSGYTPTQVSGSGTDEGIDS-----N 878
QY 1073 GYNILHRSGSVLVK-----ADGGQY---YNLG-----TTTKQFLPIQLTGEK-----1111
DB 879 GTS-----TTGVIKDKDNDTIDSGFYKPTYNLGDYVWEDTNKN--GVQDKDEKGISGVTV 931
QY 1112 --KQNEGFVK-----GNDGNYFYVDLAGNVAKNTF-----1140
DB 932 TLKDENDKVLKTVTTDENGKYQFTDLNNGTYKVFETPSGYTPTSVTSGNDTEKDSNGLT 991
QY 1141 ----IED-----SVGN--WYFFQDQGRKVENKHFVDVDSYGERGYTFEFL 1178
DB 992 TTGVIKDADNMTLDSGFYKTPKYSLGDYVWYDYSNKGQKQ-----DSTEKGIKDKVIL 1044
QY 1179 KNQYSFRGGLVQTD--NGTYFEDN 1200
DB 1045 LNEKGEVIGITKTDENGKIRFDN 1067
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 12, 2002, 07:55:23 ; Search time 34.89 Seconds

(without alignments)

1728.664 Million cell updates/sec

Title: US-09-995-749A-2_COPY_972_1514

Perfect score: 2852

Sequence: 1 QTYRQPELLANDIDNSNPV.....AIQLHKAQLQVMDWVDPQ 543

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : A_Geneseq_032802.*

1:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
2:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
3:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
4:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
5:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
6:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
7:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
8:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
9:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
10:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
11:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
12:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
13:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
14:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
15:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
16:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
17:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
18:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
19:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
20:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
21:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
22:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1532.5	53.7	2057	21	AA10667
2	1406	49.3	1592	14	AA13225
3	1288.5	45.2	1577	17	AA191047
4	169	5.9	501	21	AA14822
5	167.5	5.9	782	19	AA157436
6	165.5	5.8	750	18	AA123600
7	165.5	5.8	750	18	AA123600
8	164	5.8	501	21	AA14821
9	159.5	5.6	574	19	AA157434
10	159.5	5.6	483	14	AA131362
11	159.5	5.6	483	14	AA134289

12	159.5	5.6	483	14	AA134290
13	159	5.6	531	17	AA198007
14	158.5	5.6	483	14	AA131363
15	158.5	5.6	483	14	AA134288
16	158.5	5.6	483	15	AA158000
17	158.5	5.6	487	19	AA180215
18	158	5.6	560	18	AA123602
19	158	5.5	560	18	AA123602
20	157.5	5.5	483	21	AA199772
21	157.5	5.5	485	19	AA179904
22	157.5	5.5	516	18	AA111326
23	157.5	5.5	1349	22	AA134402
24	157.5	5.5	1349	22	AA137544
25	156.5	5.5	482	18	AA108192
26	156.5	5.5	482	18	AA108195
27	156.5	5.5	482	18	AA108196
28	156.5	5.5	482	18	AA108198
29	156.5	5.5	482	18	AA108199
30	156.5	5.5	482	18	AA108200
31	156.5	5.5	482	20	AA107390
32	156.5	5.5	483	12	AA10577
33	156.5	5.5	483	12	AA10577
34	156.5	5.5	483	12	AA10578
35	156.5	5.5	483	12	AA10579
36	156.5	5.5	483	12	AA10580
37	156.5	5.5	483	12	AA10581
38	156.5	5.5	483	12	AA10582
39	156.5	5.5	483	15	AA157993
40	156.5	5.5	483	15	AA157995
41	156.5	5.5	483	15	AA157999
42	156.5	5.5	483	15	AA158005
43	156.5	5.5	483	15	AA158007
44	156.5	5.5	483	15	AA157985
45	156.5	5.5	483	15	AA157992
46	156.5	5.5	483	15	AA158725
47	156.5	5.5	483	15	AA158726
48	156.5	5.5	483	15	AA158727
49	156.5	5.5	483	15	AA158728
50	156.5	5.5	483	15	AA158729
51	156.5	5.5	483	15	AA158730
52	156.5	5.5	483	15	AA158731
53	156.5	5.5	483	15	AA158732
54	156.5	5.5	483	15	AA158733
55	156.5	5.5	483	15	AA158734
56	156.5	5.5	483	15	AA158735
57	156.5	5.5	483	15	AA158736
58	156.5	5.5	483	15	AA158737
59	156.5	5.5	483	15	AA158738
60	156.5	5.5	483	15	AA158739
61	156.5	5.5	483	15	AA158740
62	156.5	5.5	483	15	AA158741
63	156.5	5.5	483	15	AA158742
64	156.5	5.5	483	15	AA158743
65	156.5	5.5	483	15	AA158744
66	156.5	5.5	483	15	AA158745
67	156.5	5.5	483	15	AA158746
68	156.5	5.5	483	15	AA158747
69	156.5	5.5	483	15	AA158748
70	156.5	5.5	483	15	AA158749
71	156.5	5.5	483	15	AA158750
72	156.5	5.5	483	15	AA158751
73	156.5	5.5	483	15	AA158752
74	156.5	5.5	483	15	AA158753
75	156.5	5.5	483	15	AA158754
76	156.5	5.5	483	15	AA158755
77	156.5	5.5	483	15	AA158756
78	156.5	5.5	483	15	AA158757
79	156.5	5.5	483	15	AA158758
80	156.5	5.5	483	15	AA158759
81	156.5	5.5	483	15	AA158760
82	156.5	5.5	483	15	AA158761
83	156.5	5.5	483	15	AA158762
84	156.5	5.5	483	15	AA158763

Temp-sensitive B.1	Alpha-amylase
PeB signal-Termin	Alpha-amylase
B.licheniformis al	Alpha-amylase
B.licheniformis al	Alpha-amylase
B.licheniformis al	Alpha-amylase
M197L alpha-amyl	Alpha-amylase
A4 form alpha-amyl	Alpha-amylase
Alpha-amylase-cell	Alpha-amylase
Alpha-amylase-cell	Alpha-amylase
Bacillus lichenifo	Alpha-amylase
Liquefied alkaline	Alpha-amylase
Staphylococcus aur	Alpha-amylase
Staphylococcus aur	Alpha-amylase
B.licheniformis mu	Alpha-amylase
B.licheniformis mu	Alpha-amylase
B.licheniformis mu	Alpha-amylase
B.licheniformis mu	Alpha-amylase
B.licheniformis mu	Alpha-amylase
B.licheniformis mu	Alpha-amylase
Mutant chimeric te	Alpha-amylase
Mutant alpha-amyla	Alpha-amylase
Mutant alpha-amyla	Alpha-amylase
Mutant alpha-amyla	Alpha-amylase
Mutant alpha-amyla	Alpha-amylase
M138X alpha-amyl	Alpha-amylase
M15X alpha-amylase	Alpha-amylase
M8A alpha-amylase	Alpha-amylase
M38A alpha-amyl	Alpha-amylase
M197I alpha-amyl	Alpha-amylase
Wild type alpha-am	Alpha-amylase
M15L alpha-amylase	Alpha-amylase
M15A alpha-amylase	Alpha-amylase
M15R alpha-amylase	Alpha-amylase
M15N alpha-amylase	Alpha-amylase
M15D alpha-amylase	Alpha-amylase
M15K alpha-amylase	Alpha-amylase
M138H alpha-amyl	Alpha-amylase
M15P alpha-amylase	Alpha-amylase
M15S alpha-amylase	Alpha-amylase
M15T alpha-amylase	Alpha-amylase
M15V alpha-amylase	Alpha-amylase
M15L alpha-amylase	Alpha-amylase
M15Q alpha-amylase	Alpha-amylase
M15R alpha-amylase	Alpha-amylase
M15G alpha-amylase	Alpha-amylase
M15I alpha-amylase	Alpha-amylase
M15F alpha-amylase	Alpha-amylase
M15W alpha-amylase	Alpha-amylase
M15Y alpha-amylase	Alpha-amylase
M138F alpha-amyl	Alpha-amylase
M138Y alpha-amyl	Alpha-amylase
M138A alpha-amyl	Alpha-amylase
M138L alpha-amyl	Alpha-amylase
Bacillus lichenifo	Alpha-amylase
Bacillus lichenifo	Alpha-amylase
Wild type mature a	Alpha-amylase
Alpha-amylase S187	Alpha-amylase
Alpha-amylase H133	Alpha-amylase
Alpha-amylase H133	Alpha-amylase
Alpha-amylase H133	Alpha-amylase
Alpha-amylase H133	Alpha-amylase
Alpha-amylase V128	Alpha-amylase
B. licheniformis m	Alpha-amylase
Alpha-amylase H133	Alpha-amylase
Alpha-amylase H133	Alpha-amylase
Alpha-amylase N104	Alpha-amylase
Alpha-amylase N104	Alpha-amylase

DR	WPI; 1993-079449/10.
DR	N-PSDB; NAO37760.
xx	
PT	DNA sequence glucosyl:transferase-I - comprises Streptococcus
PT	sobrinus DNA sequence with at least one nucleotide added or
xx	deleted
xx	
PS	Claim 13; Page 15; 29pp; Japanese.
xx	
CC	The DNA sequence from Streptococcus sobrinus strain 6715 encodes
CC	glucosyltransferase-I (and mutants). The DNA was obtd. by treating
CC	S. sobrinus 6715 with mutanolysin, extracting the chromosomal DNA,
CC	partially digesting with Sau3AI and fractionating on agarose gel.
CC	The 3.5 kbp fragment was ligated into pUC18 and E. coli JM109
CC	transformed with it. A gr-I expressing clone was isolated and
CC	sequenced. The clone may be used in the development of a drug for
CC	dental caries.
xx	
xx	Sequence 1592 AA:
SO	

Query Match 49.3%; Score 1406; DB 14; Length 1592;
Best Local Similarity 52.6%; Pred. No. 2.4e-90;
Matches 288; Conservative 83; Mismatches 153; Indels 24; Gaps

QY	5	GOEFLANDINSNPVVQAROLANWLYLLAFNGFITTANNDOANFDSVRVDAPONDIDALAN	64
Db	399	gydfllandvsnppvqaeqlnwlhyllnfgsiyandadafdsirvdaednvdadqiq	458
QY	65	IADQYFNAAYGMD-SDAVSNKHNILLEDKNHADPEFNKICNPQLPMDOTIKNSLHGLS	123
Db	459	issdyikaaygidkknknanhvisveawsdndcpylhddgdnlmmndknfrlsm,wsia	518
QY	124	DARN-RWGLDAIVHQSLADRENSSTENVIPNTSFVRADHNNSDQIQNAIR-DVTKDY	181
Db	519	kptdvrsqnlplhmslvrdvredvretvpsysfaraandsevgdiirdiikaeinpsf	578
QY	182	-HPTFDEQKGDIAITQDQNSVVKKNLYNIPASXAILLTFNKDTTPRVYVGLYDVGGO	240
Db	579	gysafqeeldqafkiynedliksekkythyvnpvsltlltkgsiprvyvgdmftddgq	638
QY	241	YMEHOTRYDTLTLNLKSRVYKVYAGQSMOTMSVGGNNILTSVRKVGKAMATDTGTDE	300
Db	639	ymanktvnydaieellkarmkyvaggamngqi-gngeiltsrvrygkalkgsdkgdat	697
QY	301	TRTQIGVYVSTNPNLKGLVNDKVV-LHMGAAHKNQOYRAAVLITTDGVINTSDQAP-	358
Db	698	trtsqvgvnmrgnpnfsl--dgkvva lnmgaa bancyralmwstkgdvatyetdadask	755
QY	359	---VAMTDENGDLVLSHNLVYNGKEEADTAVOGYANPDVSGYLAVWPVPGASDNQDART	415
Db	756	aglvkrtdengylyfnddl-----xgvannpqvgflqvwvvpvgaaddqdlrv	803
QY	416	APTEKNSGNSAYRTNAADPSNVIFEAFSFEVYTPKESERANVRIAQNADFPASIGFNS	475
Db	804	aasdtastdqtshqdaamdsvrmfegfsnfqsfakeeytnviannvdkfvswwgid	863
QY	476	FEMAPQNTSSKDRFTFLDSTIDNGYAFTRDYDLGMSBPNKYGTDEDLERNAIQALHKAQLOV	535
Db	864	femapqyvsstdgqfsvlqngyaftrdrydlgmskankygtadqivkaikalhaqklyv	923
QY	536	MADWVPDQ 543	
Db	924	madwvpdq 931	

RESULT 3
AAR91047
ID AAR91047 standard; Protein; 1577 AA.
XX
AC AAR91047;

DT 22-MAY-1996 (first entry)

XX	Alpha-D-glucosyltransferase.
DE	
XX	
KW	Alpha-D-glucosyltransferase; primer-independent; soluble glucan;
KW	sucrose; transgenic plant; cloning; Escherichia coli;
KW	phage lambda-CL3; vector; plasmid pGS501; plasmid pSGS502;
KW	gene transfer; crop improvement; storage carbohydrate; pasture;
KW	feedstuff; senescence; dextran; binder; food; pharmaceutical.
XX	
OS	Streptococcus salivarius strain ATCC 25975.
XX	
PN	WO9606173-A1.
XX	
PD	29-FEB-1996.
XX	
XX	24-AUG-1995; 95WO-AU00527.
PF	
XX	
PR	24-AUG-1994; 94AU-0007643.
XX	
PA	(GIFP/) GIFFARD P M.
PA	(JACO/) JACQUES N A.
PA	(SIMP/) SIMPSON C L.
XX	
PI	Giffard PM, Jacques NA, Simpson CL;
XX	
DR	WPI; 1996-151376/15.
DR	N-PSDB; AAT13139.
XX	
PT	Plants contg. new bacterial DNA encoding glucosyl transferase
PT	activity - retain higher levels of stored carbohydrate(s) in a form
PT	readily digestible by ruminants
XX	
PS	Claim 4; Page 16-20; 31pp; English.
XX	
CC	The sequence represents an alpha-D-glucosyltransferase from
CC	Streptococcus salivarius. The enzyme is primer-independent, and
CC	produces soluble glucan from sucrose. A gene encoding the enzyme
CC	may be cloned and expressed in Escherichia coli using a subclone
CC	of phage lambda-CL3, e.g. plasmid pGS501 or plasmid pGS502. The
CC	DNA may also be expressed in a transgenic plant, to improve the
CC	level of stored carbohydrate in a pasture plant which normally
CC	contains low levels, or to prevent degradation of stored carbohydrate
CC	during plant senescence. Dextran may be isolated from the plant, for
CC	use as a food binder or pharmaceutical additive. Primer independence
CC	ensures that the enzyme will be functional in plants. The glucan is
CC	poorly degraded in plants but easily degraded by bacteria in the rumen
CC	of grazing livestock.
XX	
SQ	Sequence 1577 AA;

Query Match 45.2%; Score 1288.5; DB 17; Length 1577;
Best Local Similarity 48.3%; Pred. No. 4.6e-82;
Matches 265; Conservative 94; Mismatches 163; Indels 27; Gaps

QY	5	QGEFLANDIDNSPVVQARQLNWLXYLLNFGCTITANNQANFDSVRVDAPNIDQADLWN	64
Db			
510	gyellandidsnnpvqaqelnwlyimigslngdpasufgvrivadnvdadliq	569	
QY	65	IADYFNAYGM-DSDAVSNKHNILLEDWNHADPEYFNKIGNPQLTMDTIKNS-LNHGL	122
Db			
570	lasdyfkeyryvadneanaiahslsleavsyndghykdtkgaqidsnpl-etlittfl	629	
QY	123	SDATNRWGLDAIVHQSILADRENNSTENVIPNPKSVFRAHDNNSQOIQNAI-RDVTGK-D	180
Db			
630	rksngysrslvritnslnnrseqkhtprdanyfvrhadseqvavianlskqnpktd	689	
QY	181	YHFTFEDEQKGDAYIQDQNSVFKKYNLYNPASVAILTNKDTIPRVYVYGDLYTDGQO	240
Db			
690	gfiftmdelkqafeiynadiakadkkytgynpaayatltnkdsitrvyysgdllfddgq	749	
QY	241	YMEHQTRYDTLNLKSRVYKVYVAGGSGQTSVGGNNNLTSRVYKGKAMPATDTGTD	300

Db 750 ynaekspynaidallrarikyvaggdmkvtklog-yeimssvryvgkaeeanglgtae 808
 QY 301 TRTQIGVVVSNPMLKLGVDKVLHMGAAHKNQYRAAVLTITTDGVTNYSQDQAP-- 358
 Db 809 trnqgmvlvtanrpdmkigandrlvvnmgaaaknqayrplllskstglatytkdsdvpaq 868
 QY 359 -VAMTDENDLILSSHNLVVNGKEADTAQGYANPDYSGYLAIVVPVGVASNDQARAP 417
 Db 869 lrvytdnqnlftadd-----faghnstvevsgylavvvpvgasenqdarcka 916
 QY 418 STEKNSGNSAYRTNAFDSNVIFAPSN---PVYPTKESERANVRIAQNAOFFASLGFT 474
 Db 917 sstk-kgeqvessaaalqsvayegfsnfqdvktpsqytnrv---iaqnaklfkewgft 972
 QY 475 STEMAPQYNSKDRFLDSTINDGYAFTRYDLGMSPEKNKYGTDEDLRNAIQALHKAGLQ 534
 Db 973 sfefapqvssqdgftfldsiengyafedrydiansknknkyslklmdalralhaegis 1032
 QY 535 VMADWVDPQ 543
 Db 1033 atadwvpdq 1041

RESULT 4
 ID AAB14822 standard; protein: 501 AA.
 XX AAB14822;

AC AAB14822;
 DT 19-DEC-2000 (first entry)
 XX

DE Bacillus sp. liquefying alpha-amylase #2.

XX Liquefying alpha-amylase; detergent; starch industry; brewing industry;
 KW pharmaceutical industry; food industry; fibre industry.

XX Bacillus sp.

XX EP1022334-A2.

PN 26-JUL-2000.

XX 20-DEC-1999; 99EP-0125399.

XX 21-DEC-1998; 98JP-0362487.

PR 21-DEC-1998; 98JP-0362488.

XX (KAOS) KAO CORP.

PI Hagihara H, Kitayama K, Hayashi Y, Igarashi K, Endo K, Ozaki K;

XX WPI: 2000-516014/47.

DR N-PSDB; AAA70314.

XX New liquefying alkaline amylase, useful in a detergent composition
 PT comprises a residual activity of not less than 70 % when treated under
 PT specific conditions .

PS Claim 4; Page 17-19; 34pp; English.

XX The present sequence is a novel alpha-amylase from Bacillus sp. The
 CC invention concerns the isolation of two new liquefying alpha-amylases
 CC from Bacillus sp. strains KSM-K36 and KSM-K38, designated K36 and K38, of
 CC which this protein is one. The present alpha-amylases are chelating-agent
 CC resistant and thus are useful in the starch, brewing, fibre,
 CC pharmaceutical and food industries, and especially as components of
 CC detergents.

XX Sequence 501 AA;

Query Match 5.9%; Score 169; DB 21; Length 501;
 Best Local Similarity 21.6%; Pred. No. 0.00081;

Matches 75; Conservative 44; Mismatches 106; Indels 122; Gaps 15;
 QY 7 BELLANDIDNSNPVVOAQLNWLIIYLLNFGTITANNQDQFDSVRVDAPONIDADLNMIA 66
 Db 215 dylgnsidfshpevqdelkqgswf-----tdeldldgyrlda----- 253
 QY 67 QDYFNAAVGMDSDAVSNKHIINI--LEDW-----NHADPEYF-----NKIGNPOLNDD 112
 Db 254 -----ixhipfwytswvrhgrneadqdlfvvgcywkdvdvgalefyld 297
 QY 113 TIKNSLNHGLSDATNRWGLDAIVHQSILADRENSTENVVIP-NYSFVRAHNNNSQDOIQN 171
 Db 298 m-----nw-----emslfd-----vplnyfyfraqsggsydmrn 327
 QY 172 AIRD--VTKQYHTTFEDEQ-----KCIDAYIODQNSTVKKYNLYNIPASYAILTNKD 224
 Db 328 ilrgslveabpmhavtfvdnhdtpggesleswvad-----wfkplayatiltreg 377
 QY 225 TIPRYVYGDLYTDGQYMEHQTRYDYDTLTNLLKSRVKYVAG-----GQSMQTM 272
 Db 378 gypnvfygdy---gipndnisakmdidelldarqnyaygtqhdhdyfhdwvvgwtregs 434
 QY 273 SVGGNNILITSVRYGKGA-----MTATDTGTDTRTQIGVVVS 311
 Db 435 srpnsqglatimsngpggskwmyvgrnagqtwtldtgnngasvtin 481

RESULT 5
 ID AAW57436 standard; protein: 782 AA.

XX AAW57436;
 AC AAW57436;

DT 01-SEP-1998 (first entry)

XX Termamyl-linker-CBD fusion sequence expressed in B. subtilis clone.

DE Starch; liquefaction; sweetener; enzyme hybrid; endoglucanase; enzyme;
 KW cellulose binding domain; CBD; starch processing; alpha-amylase;
 KW saccharification; Termamyl-linker-CBD fusion.

XX Bacillus sp.

OS Clostridium stercorarium.

XX WO9816633-A1.

PN 23-APR-1998.

XX 13-OCT-1997; 97WO-DK00448.

XX 11-OCT-1996; 96DK-0001130.

XX (NOVO) NOVO-NORDISK AS.

XX Bisgardfrantzen H, Bjornvad M, Pedersen S, Schulein M;

XX WPI: 1998-251283/22.

DR N-PSDB; AAV29672.

XX Liquefaction of starch for, e.g. production of sweeteners .
 PT comprises use of enzyme hybrids including cellulose binding domain
 PT for starch

PS Example 4; Pages 60-63; 83pp; English.

XX This represents the protein sequence of the termamyl-linker-CBD fusion
 CC construct expressed in B. subtilis clone MXYNCBD2. This is an enzyme
 CC hybrid which can be used for the liquefaction of starch. The enzyme
 CC hybrids contain amino acid sequences of alpha-amylase linked to a
 CC cellulose binding domain (CBD). The CBD is selected from the Bacillus
 CC endoglucanase or the Clostridium xylanase A sequence. The starch is
 CC liquefied by treating, in aqueous medium, with such an enzyme hybrid. A
 CC recombinant expression vector comprising a construct containing isolated


```

Db 215 dylgsnidfshpevqeelkdwgsf-----tdeildgyrida----- 253
QY 67 QDYFNAAYGSDDAVSNKHNII--LSDW-----NHADPEYF-----NKIGNPQLTMD 112
Db 254 -----ikhipfytadwrhqrseadqdlfvvgewkddvgaletfyle 297
QY 113 TIKNSLNHGLSDATNRWGLDAIVHQSLADRENNSTENVVYP-NYSFVRAHDNNSQDQION 171
Db 298 m-----nw-----emsldf-----vpinyfyraskgqgsydmrn 327
QY 172 AIRD--VTGKDYHTFTFEDEQ-----KGDAYIQDQNSVKKYNLYNPASVAILLTNDK 224
Db 328 lfrgsiveahphtavfvdndhtqpgesleswad-----wfkplayatiltreg 377
QY 225 TIPRVYGYDLYTDGQYMEHQTRYDTLNLLKSRVYKVVAGQS--MOTMSVGGNNILT 282
Db 378 gypnvfygdy---gipndnisakkdmideildarqnyagtqdyfhdwdivg----- 428
QY 283 SVRYGKAMTATDTGTDETRTQIGVYVSNTPNKLGVNDKVVLRHGAHKNQOYRAAVL 342
Db 429 -----wtregtssrpsnglatimsngp-----ggskmwyvgqghagq----- 465
QY 343 TTTDGVINVTSDQAGPAMTDENG 367
Db 466 twcd-----ltgnhaasvti---ngd 483

```

RESULT 9

```

AAW57434
ID AAW57434 standard; Protein; 574 AA.
XX
AC AAW57434;
XX
DT 01-SEP-1998 (first entry)
XX
DE Termamyl-1-linker-CBD fusion sequence of the plasmid pMB492.
XX
KW Starch; liquefaction; sweetener; enzyme hybrid; endoglucanase; enzyme;
KW cellulose binding domain; CBD; starch processing; alpha-amylase;
KW saccharification; Termamyl-1-linker-CBD fusion.
XX
OS Bacillus sp.
XX
OS Bacillus agaradherens.
XX
PN WO9816633-A1.
XX
PD 23-APR-1998.
XX
PF 13-OCT-1997; 97WO-DK00448.
XX
PR 11-OCT-1996; 96DK-0001130.
XX
PA (NOVO ) NOVO-NORDISK AS.
XX
PI Bisgardfrantzen H, Bjornvad M, Pedersen S, Schulein M;
XX
XX WPI; 1998-251283/22.
XX
DR N-PSDB; AAV29669.
XX
PT Liquefaction of starch for, e.g. production of sweeteners -
PT comprises use of enzyme hybrids including cellulose binding domain
PT for starch
XX
PS Claim 23; Pages 55-57; 83pp; English.
XX
CC This represents the protein sequence of the termamyl-1-linker-CBD fusion
CC construct of the plasmid pMB492. This is an enzyme hybrid which can
CC be used for the liquefaction of starch. The enzyme hybrids contain
CC amino acid sequences of alpha-amylase linked to a cellulose binding
CC domain (CBD). The CBD is selected from the Bacillus agaradherens
CC endoglucanase sequence. The starch is liquefied by treating, in aqueous
CC medium, with such an enzyme hybrid. A recombinant expression vector

```

```

CC comprising a construct containing isolated DNA encoding enzyme hybrids
CC with amylolytic activity, promoter and stop signals can be used to
CC transform host cells for the production of the recombinant enzyme
CC hybrids. The enzyme hybrids are useful in industrial starch processing
CC especially for the production of sweeteners. Hybrid enzymes have altered
CC affinity for substrate and increased activity, resulting in at least 1 of
CC reduced calcium ion dependence, reduced formation of Maillard reaction
CC products and reduced effect of alpha-amylase on subsequent
CC saccharification.
XX
SQ Sequence 574 AA;

```

```

Query Match 5.6%; Score 161; DB 19; Length 574;
Best Local Similarity 20.2%; Pred. No. 0.0036;
Matches 84; Conservative 61; Mismatches 152; Indels 118; Gaps 19;

QY 7 EFLANDIDNSNPVQAQELNWLNYLLNFGTITANNQANFDSVRVDA PDNIDADLMNIA 66
Db 223 dylmyadiidypdvaaekrw-----gtwyan--elqldgnrlav----- 262
QY 67 QDYFNAAYGMSDAVSNKHI--NILEDW-NHADPEYFNKIGNPQLTMDT IKNSL----- 118
Db 263 -----khikfsflrdwvnhvr-----ektgkemftvaeywqndlgalen 301
QY 119 -----NHGLSDATNRWGLDAIVHQSLADRENNSTENVVYPNYSFVRAHDNNSQDQION 171
Db 302 ylnktnfnhsvfdvplhyqfha-----astqgggydmrklngtvvskh----- 345
QY 172 AIRDVTGKDYHTFTFEDEQGDIDAYIQDQNSVKKYNLYNPASVAILLTNDKTI PRVY 231
Db 346 plksvtfvdnh-----dtqpg-----qslbstvqtwfk---playafiltresgtpqvfy 392
QY 232 GDLYTDGQYMEHQTRYDTLNLLKSRVYKVVAGQSMQTM SVGGNNILTSVRYKKGAM 291
Db 393 gdmvgtkgsqreipalkkkipilkarqyaygaqh---dyfdhhddivgtregdsav 448
QY 292 TATDTGTDETRTQIGVYVSNTPNKLGVNDKVVLRHGAHKNQOYRAAVLTITDGVINY 351
Db 449 a-----nsglaaltdpgp---gakr---myvgrnagetwh-----di 481
QY 352 TSDQGAPVAMTDEN--GDLYLSSHNL--VVRNGKEADTAVQGYANPDVSGYLAVVVP 404
Db 482 tgnrseppvlnsegwefhvnngsvsiyqrsdpdsgepdtppsdpgeypawdp 536

```

RESULT 10

```

AAAR31362
ID AAR31362 standard; Protein; 483 AA.
XX
AC AAR31362;
XX
DT 19-MAY-1993 (first entry)
XX
DE B.licheniformis alpha amylase Tyr 133 mutein.
XX
KW thermostability; site-directed mutagenesis.
XX
OS Bacillus licheniformis.
XX
FH Key Location/Qualifiers
FT Misc-difference 133
FT /note= "wild-type His subst.d. by Tyr"
XX
PN FR2676456-A.
XX
PD 20-NOV-1992.
XX
PF 13-MAY-1991; 91FR-0005740.
XX
PR 13-MAY-1991; 91FR-0005740.
XX
PA (INRG ) INRA INST NAT RECH AGRONOMIQUE.

```

XX PI Declerck K, Gaillardin C, Joyet P;
 XX DR WPI; 1993-020251/03.
 XX DR P-PSDB; AAQ34689.
 XX PT New heat stable forms of Bacillus licheniformis alpha-amylase -
 XX PT have specific aminoacid substitutions at positions 133 and/or 209
 XX PT for simultaneous gelation and liquefaction of starch e.g. in
 XX PT brewing
 XX PS Claim 3; Fig 1 and page 5; 23pp; French.
 XX CC Codon 133 of B.licheniformis alpha-amylase (CAC,HIS) was mutated to
 CC TAC (Tyr) by site-directed mutagenesis. The enzyme encoded by the
 CC mutant sequence is a specifically claimed example of an alpha-
 CC amylase in which His 133 is replaced by a more hydrophobic amino
 CC acid. The Tyr133 mutant had a half-life at 90 deg.C of 8 mins, c.f.
 CC 5 mins for the wild-type enzyme. Substitution of Ala209 by a more
 CC hydrophobic amino acid also enhances the enzyme's thermostability.
 CC When the enzyme carries mutations at both positions 133 and 209 it
 CC has a half-life at 90 deg.C that is 9-10 times greater than the
 CC native enzyme. The mutant alpha-amylases are useful in the brewing,
 CC papermaking and textile industries. See also AAR31363.
 XX SQ Sequence 483 AA;

Query Match 5.6%; Score 159.5; DB 14; Length 483;
 Best Local Similarity 22.9%; Pred. No. 0.0036;
 Matches 78; Conservative 39; Mismatches 119; Indels 105; Gaps 14;
 QY 7 EFLANDIDNSNPVQAEQLNWLNYLLNFGTITANNQANFDSVRVDAPDNIDALMNTA 66
 Db 194 dYlmYadIdYdhpDvaeeIKrw-----gtwyan--elqldgfrIdav----- 233
 QY 67 QDYFNAAYGMDSDAVSNKHI--NILEDW-NHADPEYFNKIGNPQLTMDDTIKNSL----- 118
 Db 234 -----khikfsfirdwvnhvr-----ektgkemtvaeywqndlgalen 272
 QY 119 -----NHGLSDATNRWGLDAIVHOSLADRENNSTENVIPYSFVRAIDNNSSQDQIQN 171
 Db 273 Ylnktfnhslfdvplhyqfha-----astggggydmrkilngtvvskh----- 316
 QY 172 AIRDVTGKDYHTFTFEDEQKIDAYIQDQNSTVKKNLYNIPASYAILLTNKDTPRYY 231
 Db 317 plksvtfvDnh-----dtqpg-----qlestvqtWfk---playafiltresgypqvfy 363
 QY 232 GDLYTDGGQYMEHQTRYDYDLTNLLKSRVKYVAGGQ-----SMQTMMSGVGN 279
 Db 364 gdmYgtkgdsqreipalkhkiepikarkqaygaqhdYfdhhdvgtwregdssvansg 423
 QY 280 ILTSVRYGKGA-----MTATDTGTDETRTQIGGVVSN 312
 Db 424 iaalitdpggagrmyvgrqnagetwhdltgrseppvins 464

RESULT 11
 ID AAR34289
 XX AAR34289 standard; Protein; 483 AA.
 XX AC AAR34289;
 XX DT 19-MAY-1993 (first entry)
 XX DE B.licheniformis alpha amylase mutein with substn. at position 133.
 XX KW thermostability; site-directed mutagenesis.
 XX OS Bacillus licheniformis.
 XX FH Key Location/Qualifiers
 FT Misc-difference 133

FT FT /label= Phe, Gln, Glu, Lys, Leu, Tyr
 FT FT /note= "substn. of wild-type His133 by a more
 FT FT hydrophobic amino acid, esp. Tyr,
 FT FT increases the thermostability of alpha-
 XX XX amylase"
 PN PR2676456-A.
 XX 20-NOV-1992.
 XX 13-MAY-1991; 91FER-0005740.
 XX 13-MAY-1991; 91FER-0005740.
 PA (INRG) INRA INST NAT RECH AGRONOMIQUE.
 PI Declerck K, Gaillardin C, Joyet P;
 DR WPI; 1993-020251/03.
 XX New heat stable forms of Bacillus licheniformis alpha-amylase -
 XX PT have specific aminoacid substitutions at positions 133 and/or 209
 XX PT for simultaneous gelation and liquefaction of starch e.g. in
 XX PT brewing
 XX PS Claim 1; Fig 1; 23pp; French.

XX CC A series of alpha-amylase mutants was prepared by site-directed
 CC mutagenesis at amino acid position 133. Substitution of wild-type
 CC His by Pro, Ser, Gly or Ala (all are less hydrophobic than His)
 CC resulted in proteins with less residual activity after heating than
 CC the native enzyme. Replacement of His by Gln, Lys, Glu, Leu, Phe
 CC and Tyr (increasingly more hydrophobic than His) resulted in
 CC proteins with a greater residual activity after heating than His.
 CC The Tyr133 mutant had a half-life at 90 deg.C of 8 mins, c.f.
 CC 5 mins for the wild-type enzyme. Substitution of Ala209 by a more
 CC hydrophobic amino acid also enhances the enzyme's thermostability.
 CC When the enzyme carries mutations at both positions 133 and 209 it
 CC has a half-life at 90 deg.C that is 9-10 times greater than the
 CC native enzyme. The mutant alpha-amylases are useful in the brewing,
 CC papermaking and textile industries.
 CC See also AAR31362-3, AAR34288 and AAR34290.

XX SQ Sequence 483 AA;

Query Match 5.6%; Score 159.5; DB 14; Length 483;
 Best Local Similarity 22.9%; Pred. No. 0.0036;
 Matches 78; Conservative 39; Mismatches 119; Indels 105; Gaps 14;
 QY 7 EFLANDIDNSNPVQAEQLNWLNYLLNFGTITANNQANFDSVRVDAPDNIDALMNTA 66
 Db 194 dYlmYadIdYdhpDvaeeIKrw-----gtwyan--elqldgfrIdav----- 233
 QY 67 QDYFNAAYGMDSDAVSNKHI--NILEDW-NHADPEYFNKIGNPQLTMDDTIKNSL----- 118
 Db 234 -----khikfsfirdwvnhvr-----ektgkemtvaeywqndlgalen 272
 QY 119 -----NHGLSDATNRWGLDAIVHOSLADRENNSTENVIPYSFVRAIDNNSSQDQIQN 171
 Db 273 Ylnktfnhslfdvplhyqfha-----astggggydmrkilngtvvskh----- 316
 QY 172 AIRDVTGKDYHTFTFEDEQKIDAYIQDQNSTVKKNLYNIPASYAILLTNKDTPRYY 231
 Db 317 plksvtfvDnh-----dtqpg-----qlestvqtWfk---playafiltresgypqvfy 363
 QY 232 GDLYTDGGQYMEHQTRYDYDLTNLLKSRVKYVAGGQ-----SMQTMMSGVGN 279
 Db 364 gdmYgtkgdsqreipalkhkiepikarkqaygaqhdYfdhhdvgtwregdssvansg 423
 QY 280 ILTSVRYGKGA-----MTATDTGTDETRTQIGGVVSN 312
 Db 424 iaalitdpggagrmyvgrqnagetwhdltgrseppvins 464

QY 7 EFLIANDIDNSNPVQAEQLNWLYLLNFGTITANNQANFDSVVDAPDNIDADLNIA 66
 Db 194 dyllyadiydhpdvaaekrw-----gtwan--elqldgfrldav----- 233
 QY 67 QDYFNAAYGMSDAVSNKHI--NILEDW-NHADPEYFNKIGNPQLTMDDTIKNSL----- 118
 Db 234 -----khikfsfrldvnhvr-----ektgkemftvaeywqndlgalen 272
 QY 119 -----NHGLSDATNRWGLDAIVHQSADRENSTENVVPIPNVSVFRAHDNNSQDIQON 171
 Db 273 ylnktnfnhsvfdvplhyqfha-----astqgggydmrklngtvvskh----- 316
 QY 172 AIRDVTGDKDYHTFFEDQKGDIDAYIQDNSTVKYKLYNIPASVAILLNKDTIPRVY 231
 Db 317 plksvtfvdnh-----dtqpg-----qslstvtqfwk---playafiltresgypqvfy 363
 QY 232 GDLTYDGGQYMEHQTRYDITNLKSKRVKVVAGGQ-----SMQTMVSGGN 279
 Db 364 gdmvgtkgdsqreipalkhkiepalkarkayagqahdyfdhhdvgtregdssvansg 423
 QY 280 ILTSVRVYKGA-----MTATDTGDTETRTQIGVYVSN 312
 Db 424 iaaltdpggagrmygrqnagetwhdtgnrseppvins 464

RESULT 17

AAW80215
 ID AAW80215 standard; Protein; 487 AA.

AC AAW80215;

XX

DT 04-JAN-1999 (first entry)

De A4 form alpha-amylase variant M197L.

XX Bacillus amyloliquefaciens; transcriptional terminator;

KW alpha-amylase protein; variant.

KW

OS Synthetic.

OS Bacillus licheniformis.

XX

Key Location/Qualifiers

FT Misc-difference 201

FT /label= M201L

FT /note= "wild type Met197 replaced with Leu"

XX US5824532-A.

XX 20-OCT-1998.

XX 06-JUN-1995;

XX 95US-0468220.

XX 10-FEB-1994;

XX 94US-0194664.

XX 11-FEB-1993;

XX 93US-0016395.

XX 06-JUN-1995;

XX 95US-0468220.

XX (GENV) GENENCOR INT INC.

XX Barnett CC, Mitchinson C, Power SD, Requaad CA;

XX WPI: 1998-582560/49.

XX DNA encoding mutant and variant alpha-amylase proteins - of Bacillus

XX licheniformis, useful for producing recombinant alpha-amylase

XX proteins

XX Example 3; Page -; 56pp; English.

XX

CC The present sequence represents a mutant Bacillus licheniformis

CC alpha-amylase protein. M197T and M197T (A4 form) mutants have

CC enhanced stability in detergent formulations. M15X mutants (where X = D,

CC N, T or L) have high specific activities and enhanced starch

CC liquefaction performance at pH 5.5. M197T/M138X double mutants

CC

CC

CC

CC

CC (where X = F or Y) have enhanced resistance to chloramine-T oxidation.
 CC note: the present sequence does not appear in the specification; it was
 CC created using information provided.

XX Sequence 487 AA;

SO

Query Match

Best Local Similarity 5.6%; Score 158.5; DB 19; Length 487;

Matches 78; Conservative 39; Mismatches 119; Indels 105; Gaps 14;

QY 7 EFLIANDIDNSNPVQAEQLNWLYLLNFGTITANNQANFDSVVDAPDNIDADLNIA 66

Db 198 dyllyadiydhpdvaaekrw-----gtwan--elqldgfrldav----- 237

QY 67 QDYFNAAYGMSDAVSNKHI--NILEDW-NHADPEYFNKIGNPQLTMDDTIKNSL----- 118

Db 238 -----khikfsfrldvnhvr-----ektgkemftvaeywqndlgalen 276

QY 119 -----NHGLSDATNRWGLDAIVHQSADRENSTENVVPIPNVSVFRAHDNNSQDIQON 171

Db 277 ylnktnfnhsvfdvplhyqfha-----astqgggydmrklngtvvskh----- 320

QY 172 AIRDVTGDKDYHTFFEDQKGDIDAYIQDNSTVKYKLYNIPASVAILLNKDTIPRVY 231

Db 321 plksvtfvdnh-----dtqpg-----qslstvtqfwk---playafiltresgypqvfy 367

QY 232 GDLTYDGGQYMEHQTRYDITNLKSKRVKVVAGGQ-----SMQTMVSGGN 279

Db 368 gdmvgtkgdsqreipalkhkiepalkarkayagqahdyfdhhdvgtregdssvansg 427

QY 280 ILTSVRVYKGA-----MTATDTGDTETRTQIGVYVSN 312

Db 428 iaaltdpggagrmygrqnagetwhdtgnrseppvins 468

RESULT 18

AAW23602

ID AAW23602 standard; Protein; 560 AA.

XX

AC AAW23602;

XX

DT 08-FEB-1998 (first entry)

De Alpha-amylase-cellulose binding domain hybrid enzyme.

XX Desizing; cellulose; fabric; enzyme hybrid; alkaline cellulase;

KW endoglucanase; Cel5A; cellulose binding domain; XynA; MBamyCSA;

KW Bacillus licheniformis; Bacillus agaradherens.

OS Chimeric - Bacillus licheniformis.

OS Chimeric - Bacillus agaradherens strain NCIMB 40482.

XX W09728243-A1.

XX 07-AUG-1997.

XX 29-JAN-1997;

XX 97WO-DK00042.

XX 29-JAN-1996;

XX 96DK-0000094.

XX (NOVO) NOVO-NORDISK AS.

XX Bjornvad ME, Cherry JR, Rasmussen MD, Vind J, Von Der Osten C;

XX WPI: 1997-402598/37.

XX N-PSDB; AAW74271.

XX

XX Cleaning of cellulosic fabrics - using an enzyme hybrid comprising a

XX sequence of a non-cellulolytic enzyme linked to a cellulose-binding

XX domain sequence

XX

XX Claim 17; Page 99-101; 72pp; English.

PS

AA999772

ID AAY99772 standard; Protein: 483 AA.

XX AAY99772;

XX AAY99772;

DT 04-SEP-2000 (first entry)

DE Bacillus licheniformis Termamyl alpha-amylase mutant KI176R-I201F+H205N.

XX

KW Bacillus; alpha-amylase; washing; textile desizing;

KW starch liquefaction; saccharification; mutagen; mutant;

XX enzyme stability; hybrid.

XX

OS Bacillus licheniformis.

OS Synthetic.

XX

FH Key Location/Qualifiers

FT Misc-difference 176

FT /note= "Wild-type Lys substituted by Arg"

FT Misc-difference 201

FT /note= "Wild-type Ile substituted by Phe"

FT Misc-difference 205

FT /note= "Wild-type His substituted by Asn"

FT WO200029560-A1.

FN 25-MAY-2000.

XX

XX 16-NOV-1999; 99WO-DK00628.

XX 16-NOV-1998; 98DK-0001495.

XX (NOVO) NOVO-NORDISK AS.

XX Svendsen A, Kjaerulf S, Bisgard-Frantzen H, Andersen C;

XX WPI; 2000-387777/33.

XX N-PSDB; AAA48483.

XX Variant of parent termamyl-like alpha amylase useful for washing,

XX textile desizing and starch liquefaction, comprising alterations in one

XX or more solvent exposed amino acid residues

XX Claim 15; Page 7; 80pp; English.

XX The present sequence is a mutant alpha amylase with increased stability

XX at acidic pH, low calcium concentration and high temperatures compared

XX to the parent enzyme. The parent enzyme is commercially available as

XX Termamyl. A variant may contain mutations in one or more solvent exposed

XX amino acid residues to increase the overall hydrophobicity of the enzyme

XX or the overall number of methyl groups in the side chains of exposed

XX residues may be increased. The mutations can be incorporated by

XX site-directed mutagenesis or by random mutagenesis. As a result of their

XX increased stability, the variants are suitable for the industrial

XX processing of starch, i.e. starch liquefaction and saccharification.

XX They may also be useful for washing, dishwashing and textile desizing.

XX Hybrid alpha-amylases comprising partial amino acid sequences derived

XX from two or more alpha-amylases have also been created in order to

XX increase enzyme stability.

XX Note: The present sequence is not shown in the specification but is

XX derived from the Bacillus licheniformis wild-type alpha-amylase sequence

XX shown in AAY99605.

XX Sequence 483 AA;

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

QY 67 QDYENAAAGMSDAVSXKHI--NILEDW-NHADPEYENKIGNPOLTFMDDTIKNSL----- 118

Db 234 -----khikfsfirdwvnhvr-----ektgkemftvaeywqndigalen 272

QY 119 -----NHGLSDATNKGDAIVHQSLADRENSTENVVPIPNYSFVRAHDNNSQDQIQN 171

Db 273 ylnktnfnhsvfdvplhyqfha-----astqgggydmrkllngtvvskh----- 316

QY 172 AIRDVTGKDYHTTFEEOQKIDAYIQDQNSTVKYKYNLYNPASAYAILLTNKTIPRVY 231

Db 317 plksvtfvdnh-----dtqpg-----qslstvtqtfk---playafiltresgybpqvy 363

QY 232 GDLYTDGGQYMEHQTRYDTLTNLKSRVYVAGGQ-----SMOTMSVGGNNN 279

Db 364 gdaygtkgdsqreipaikhkkipkarkayagachdyfdhhdvgtwtregdssvansg 423

QY 280 ILFSRVYKGA-----MTATDTGTDETFRTGIGVWVSN 312

Db 424 laaltdpgpgakrmvgrqnagetwhdltgnrsepvvlns 464

RESULT 21

AAW79904

ID AAW79904 standard; protein: 485 AA.

XX AC AAW79904;

XX 18-DEC-1998 (first entry)

DE Liquefied alkaline alpha-amylase.

XX Liquefied alkaline alpha-amylase; mutation; Bacillus licheniformis;

KW detergent; bleaching agent; oxidising agent.

XX Bacillus licheniformis.

FH Key Location/Qualifiers

FT Misc-difference 202

FT /note= "can be deleted or substituted by another amino

FT acid residue such as Thr, Ile, Leu, Ala, Val

FT or Ser"

XX WO9844126-A1.

XX 08-OCT-1998.

XX 31-MAR-1998; 98WO-JP01464.

XX 31-MAR-1997; 97JP-0080299.

XX (KAOS) KAO CORP.

XX Hatada Y, Ikawa K, Ito S;

XX WPI; 1998-542707/46.

XX Bacillus derived alpha amylase having mutation at position 202

XX has optimum pH in alkaline conditions and high tolerance to

XX oxidants, useful for production of detergent compositions

XX Claim 1; Page 19-21; 42pp; Japanese.

XX The present sequence represents Bacillus licheniformis liquefied alkaline

XX alpha-amylase. The present invention describes mutated forms of the

XX liquefied alkaline alpha-amylase derived from Bacillus species KSM-API378

XX (FEW EP-3048), having the methionine residue at position 202 either

XX deleted or substituted by another amino acid, such as threonine,

XX isoleucine, leucine, alanine, valine or serine. The mutated enzyme may

XX have other mutations (such as deletion of arginine or glycine at

XX positions 181 and 182, respectively) but at least 95.2% homologous to

XX the original enzyme. The mutated enzyme has optimum pH in alkaline

XX conditions, a high alpha amylase activity, and a high and sustained

XX

XX

XX

XX

XX

Query Match 5.8%; Score 157.5; DB 21; Length 483;

Best Local Similarity 22.6%; Pred. No. 0.005;

Matches 77; Conservative 39; Mismatches 120; Indels 105; Gaps 14;

QY 7 EFLANDINSNPVQAEQUNWLYLNFCTITANNQANFSDVRDAPDNIDADLMNTA 66

Db- 194 Yimayadfydnpdvadeelkrw-----gtwyan--elqldgfridav----- 233

CC tolerance to oxidising substances. The enzyme may be used in the
 CC formulation of liquid, powder or granular detergent compositions,
 CC especially those containing bleaches and oxidants. The enzyme retains
 CC high activity in the presence of bleaches and oxidants allowing improved
 CC detergent formulations to be produced.

XX Sequence 485 AA;

Query Match 5.5%; Score 157.5; DB 19; Length 485;
 Best Local Similarity 22.7%; Pred. No. 0.0055;
 Matches 85; Conservative 49; Mismatches 123; Indels 117; Gaps 19;

Qy 7 EFLANDIDNSNFWQAEQWLYLLNFGTITANNDQANFDSVRYDAPDNIDADLNNIA 66

Db 199 dylmyadidmdhpevineirngvwytn-----tlnldgfridav----- 238

Qy 67 QDYNAAAGDSDAVSNKHI--NILEDW-NHADPEYFNKIGNPQLTMDDTIKNSL---N 119

Db 239 -----khkysytrdwthvr-----ntgkpmfavaefwkndlaaen 277

Qy 120 HGLSDATNRWGLDAIVHQSLADRENN---STENVVTPNYSFVRAHDNNSQDIQNAIRD 175

Db 278 ylnktswnhsvfdvplhynlnasnggyfdmnnli--ngsvvqkh-----piha 325

Qy 176 VTGKDYHTFTFEDQKG--IDAYIQDNSTVKKYLNIPASAILLTKDTPRVYGD 233

Db 326 vtfvdnh-----dsqgealesfvgs-----wfkplayaliltregypsvfygd 370

Qy 234 LY---TDGGQVMEHOTRYDYDTNLKLSRVKVVAGGSMQTMVGNNILTSVRYGKGA 290

Db 371 yygtpthgvpsmksk-----idplqarqtayvgtgh---dyfhdhdiigwtregdss 420

Qy 291 MTATDTGTDETRTQIGVWVSNTPNLKLGVNDKVLHMGAAHKNQYR-----AAVLT 343

Db 421 hp-----nsglatimsdgp-----ggnk---wmyvgkhkagvwrlditgnrgvtl 464

Qy 344 TTDGVINTSDQGA 357

Db 465 nadgwnftvngga 478

RESULT 22

AAW11326

ID AAW11326 standard; Protein; 516 AA.

XX AAW11326;

AC AAW11326;

XX 11-NOV-1997 (first entry)

XX Alkaline liquefying alpha-amylase.

XX Alkaline liquefying alpha-amylase; Bacillus; alpha-amylase; detergent;
 XX starch-related polysaccharide; hydrolysis; enzyme; surfactant resistance;
 XX alpha-1,4-glucosidic link; alkalophilic Bacillus; laundry detergent;
 XX dish-washing detergent; starch.

XX Bacillus species KSM-AP1378.

XX WO9700324-A1.

XX 03-JAN-1997.

XX 14-JUN-1996; 96WQ-JP01641.

XX 14-JUN-1995; 95JP-0147257.

XX (KAOS) KAO CORP.

XX Ara K, Hatada Y, Ito S, Kawai S, Ozaki K;

XX WPI; 1997-118708/11.

XX N-PSDB; AAT51339.

XX DNA encoding alkaline liquefying alpha-amylase - useful in
 PT dish-washing and laundry detergents for removal of starch dirt
 XX Claim 2; Page 23-26; 40pp; English.

XX This sequence represents an alkaline liquefying alpha-amylase.
 CC Alpha-amylase is an enzyme that acts on starch-related polysaccharides,
 CC hydrolysing the alpha-1,4-glucoside bond of the polysaccharide molecule.
 CC Alkaline liquefying alpha-amylases exhibit resistance to surfactants used
 CC in detergents, and decompose starch or starch-related polysaccharides in
 CC a highly random manner. The Bacillus species KSM-AP1378, from which this
 CC sequence was isolated, is an alkalophilic Bacillus strain. It was
 CC isolated from soil in the vicinity of the city of Tochigi. The enzyme is
 CC useful in improving the efficiency of dish-washing and laundry
 CC detergents, particularly on starch dirt.

XX Sequence 516 AA;

Query Match 5.5%; Score 157.5; DB 18; Length 516;
 Best Local Similarity 22.7%; Pred. No. 0.0055;
 Matches 85; Conservative 49; Mismatches 123; Indels 117; Gaps 19;

Qy 7 EFLANDIDNSNFWQAEQWLYLLNFGTITANNDQANFDSVRYDAPDNIDADLNNIA 66

Db 230 dylmyadidmdhpevineirngvwytn-----tlnldgfridav----- 269

Qy 67 QDYNAAAGDSDAVSNKHI--NILEDW-NHADPEYFNKIGNPQLTMDDTIKNSL---N 119

Db 270 -----khkysytrdwthvr-----ntgkpmfavaefwkndlaaen 308

Qy 120 HGLSDATNRWGLDAIVHQSLADRENN---STENVVTPNYSFVRAHDNNSQDIQNAIRD 175

Db 309 ylnktswnhsvfdvplhynlnasnggyfdmnnli--ngsvvqkh-----piha 356

Qy 176 VTGKDYHTFTFEDQKG--IDAYIQDNSTVKKYLNIPASAILLTKDTPRVYGD 233

Db 357 vtfvdnh-----dsqgealesfvgs-----wfkplayaliltregypsvfygd 401

Qy 234 LY---TDGGQVMEHOTRYDYDTNLKLSRVKVVAGGSMQTMVGNNILTSVRYGKGA 290

Db 402 yygtpthgvpsmksk-----idplqarqtayvgtgh---dyfhdhdiigwtregdss 451

Qy 291 MTATDTGTDETRTQIGVWVSNTPNLKLGVNDKVLHMGAAHKNQYR-----AAVLT 343

Db 452 hp-----nsglatimsdgp-----ggnk---wmyvgkhkagvwrlditgnrgvtl 495

Qy 344 TTDGVINTSDQGA 357

Db 496 nadgwnftvngga 509

RESULT 23

AAU34402

ID AAU34402 standard; Protein; 1349 AA.

XX AAU34402;

XX 14-FEB-2002 (first entry)

XX Staphylococcus aureus cellular proliferation protein #678.

XX Antisense; prokaryotic cellular proliferation protein;

XX antibiotic; antibacterial; drug design.

XX Staphylococcus aureus.

XX WO200170955-A2.

XX 27-SEP-2001.

XX 21-MAR-2001; 2001WO-US09180.

XX PR 21-MAR-2000; 2000US-191078P.
XX PR 23-MAY-2000; 2000US-206848P.
XX PR 26-MAY-2000; 2000US-207727P.
XX PR 23-OCT-2000; 2000US-242578P.
XX PR 27-NOV-2000; 2000US-253625P.
XX PR 22-DEC-2000; 2000US-257931P.
XX PR 16-FEB-2001; 2001US-269308P.
XX PA (ELIT-) ELITRA PHARM INC.
XX PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
XX PI Yamamoto RT, Xu HH;
XX DR WPI; 2001-611495/70.
XX DR N-PSDB; AAS52261.
XX PT New polynucleotides for the identification and development of
XX PT antibiotics, comprise sequences of antisense nucleic acids -
XX PS Example 3; Seq ID No 5898; 511pp; English.
XX CC The invention relates to antisense inhibitors of genes essential to
XX CC prokaryotic cellular proliferation, their use in identifying the
XX CC genes, their use in the discovery of novel antibiotics, the essential
XX CC genes themselves and the encoded proteins. The prokaryotes used are
XX CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
XX CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
XX CC invention is also useful for the identification of potential new targets
XX CC for antibiotic development. The antisense nucleic acids can also be used
XX CC to identify proteins used in proliferation, to express these proteins,
XX CC and to obtain antibodies capable of binding to the expressed proteins.
XX CC The proteins can be used to screen compounds in rational drug discovery
XX CC programmes. The antisense nucleic acid sequence is also useful to screen
XX CC for homologous nucleic acids which are required for cell proliferation in
XX CC a wide variety of organisms. The present sequence represents an
XX CC essential prokaryotic cellular proliferation protein.
XX CC Note: The sequence data for this patent did not form part
XX CC of the printed specification, but was obtained in electronic
XX CC format directly from WIPO at
XX CC ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 1349 AA;

Query Match 5.5%; Score 157.5; DB 22; Length 1349;
Best Local Similarity 20.0%; Pred. No. 0.02;
Matches 126; Conservative 84; Mismatches 235; Indels 185; Gaps 29;

QY 13 DIDNSNPVQAEQLNML-----YVLL-----NFGTITAN 41
DB 261 dadknnkiyqaqdylsksqitvdkvksqdyftikysdtvqvynpeditknigdkdp 320
QY 42 NDAQNFDVSRVDAPNI-----DADLNIADQYFNAAAYGMDSDAVS-NKHTNITLEDWN 93
DB 321 nngetiatakhdannlityftdyvdrnsqgmnginsiyndadtivpskn----- 372
QY 94 HADPEYFNKIGN-----PQLTMDTITKNSLHGLSDATNKGWLDATVHQSLADR 142
DB 373 --dvefnvntgttttttntaniqydpdyvne--knsigsafetvshvg-----nk 418
QY 143 EN-----NSTENVVIPNV SFYRAHDNNSQDQIQNAIRDVT----- 177
DB 419 enpgyykqtiyvpaeusltnaklkqayhssypnnigqinkdvtidikiyvpkqytlmk 478
QY 178 GKDYHFTFDEKGDVATIQ-----DQNSTVKKYKLYNIPASVAILLNKDTIPRVYVG 232
DB 479 gydvntkeltidv---tnqylqkitygdnnnsavidfg--nadsayvwmvntk----- 524
QY 233 DLYTDGQGYMEHOTRYDITLNLKSRVYVAGGQSW---QTSVSGNNILTSVRYGKG 289
DB 525 -----fqynsesptlvqmatlssstgknsvatgnaigtfnngsgagdev-----ykig 573

QY 290 AMEATWTF---GTDETRTQIG-----VVSNTPTNLKLG--VNDKVVYLMGAAHKNQYRAA 340
DB 574 nyvwedtnkngvgelgekgvgnvtvtvfdntntkvgavtkedgsyllnlpngdyrve 633
QY 341 VLTITDGVINYTSDQAPVAMTDENGDLVLSHNLVYVNGKEE----- 382
DB 634 fsnlpkgyevtpskggnceel--dsng---iss-vitvngkdnlsadlgiykpknlycdyv 688
QY 383 -ADTAVQGYANPDVSGYLAVVYVPVGASDNQDARTAPSTKNSGNSAYRTNAAFDSNVIFE 441
DB 689 wedtnkngiqddekdgisgvvtvl-kdengdvlktvttd---adgkykftdlhngnykve 744
QY 442 AFSNFVYPTKESERANVRIAQNAFFASGCTSFEMAPQYNSKDRFLDSTIDNGYAF 501
DB 745 fttpegytptvtstgsdiekdn-----gitttgi-----ngad-----nmtldsgfyk 789
QY 502 TDRYDLG---MSEPNKYGTDEDLRINAIAL 528
DB 790 tpkynlignyvwedtnkdkqkstegkigsv 819
RESULT 24
AAU37544
ID AAU37544 standard; Protein; 1349 AA.
AC AAU37544;
XX
DT 14-FEB-2002 (first entry)
XX
DE Staphylococcus aureus cellular proliferation protein #1714.
XX
KW Antisense; prokaryotic cellular proliferation protein;
XX KW antibiotic; antibacterial; drug design.
XX OS Staphylococcus aureus.
XX PN WO2001170955-A2.
XX PD 27-SEP-2001.
XX PF 21-MAR-2001; 2001WO-US09180.
XX PR 21-MAR-2000; 2000US-191078P.
XX PR 23-MAY-2000; 2000US-206848P.
XX PR 26-MAY-2000; 2000US-207727P.
XX PR 23-OCT-2000; 2000US-242578P.
XX PR 27-NOV-2000; 2000US-253625P.
XX PR 22-DEC-2000; 2000US-257931P.
XX PR 16-FEB-2001; 2001US-269308P.
XX PA (ELIT-) ELITRA PHARM INC.
XX PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
XX PI Yamamoto RT, Xu HH;
XX DR WPI; 2001-611495/70.
XX DR N-PSDB; AAS55403.
XX PT New polynucleotides for the identification and development of
XX PT antibiotics, comprise sequences of antisense nucleic acids -
XX PS Example 3; Seq ID No 13137; 511pp; English.
XX CC The invention relates to antisense inhibitors of genes essential to
XX CC prokaryotic cellular proliferation, their use in identifying the
XX CC genes, their use in the discovery of novel antibiotics, the essential
XX CC genes themselves and the encoded proteins. The prokaryotes used are
XX CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
XX CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
XX CC invention is also useful for the identification of potential new targets
XX CC for antibiotic development. The antisense nucleic acids can also be used
XX CC to identify proteins used in proliferation, to express these proteins,
XX CC and to obtain antibodies capable of binding to the expressed proteins.

CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
XX
SQ Sequence 1349 AA;

Query Match 5.3%; Score 157.5; DB 22; Length 1349;
Best Local Similarity 20.0%; Pred. No. 0.02;
Matches 126; Conservative 84; Mismatches 235; Indels 185; Gaps 29;

QY 13 DIDNSPVVQAEQNLNWL-----YVLL-----NFGTITAN 41
Db 261 dadknnkivpaqdyvlsksqitvdvksgdyftikysdvtvgyglnpeditknigdkdp 320
QY 42 NDQANFDSVRVADPNV-----DADLMIADQYFNAYGMSDAYS-NKHINILEDWN 93
Db 321 nngetiatakdhtannlityftdyvdrfnsvgmngnysiymdadtivpskn----- 372
QY 94 HADPEYFNKIGN-----PQLTMDDTIKNSLNHGLSDATNRWGLDAIVHQSLADR 142
Db 373 --dvefnvntgntttkttanlqpydyvne--knsigsafetvshv-----nk 418
QY 143 EN-----NSTENVIPNYSFVRADNNSQDQIQNAIRDVT----- 177
Db 419 enpgyykqtiyvpsenslnaklkqaybsyppnnigqinkvtdkikyqpkgytlnk 478
QY 178 GKDYHRTFEDEKGDVAYTQ-----DQNSTYKYNLYNTPASVAILLTKNDIPRYVY 232
Db 479 gydvntkeltv-----tnqyigkitygdnsavifg--nadsayvmvntk----- 524
QY 233 DLXTDGGQYMEHQTRYTYDTLTLLKSRVKYVAGGQSM---QTMVSGVGNNNILTSVRVYK 289
Db 525 -----fqylnseaptlvqmatlsstgnkavstgnalgtntnsgggagqev-----yklg 573
QY 290 AMPATDN---GTDETRTQGTG-----VVVSFPPNKLIG--VNDKVVYLMGAHKNQYRAA 340
Db 574 nywvedtnkngvqelgekgyvntvfvtdntntkvgavtkedgsvlylnlpgdyrve 633
QY 341 VLATTGQVINYTSQGPVAMTDENGDLYLSSHNLYVNGKEE----- 382
Db 634 fsnlpgyevtpskggnnee--dsng---lss-vitvngkdnlsadlgiykpkylnicdyv 688
QY 383 -ADTAVQGYANPDVSGYLAWVPVPGASDQOARTAPSTKNSGNSAYRTNAAFDSNVIFE 441
Db 689 wedtnknglqddqekglsylvtl-kdengvltkttd---adgkykftdlhngnykve 744
QY 442 AFSNVFVYTPKESERANVIAQNAADFASLGFTSFEMAPQYNSKDRTEFLDSTIDNGYAF 501
Db 745 fttpegtytptvtsgsdiekdsn-----glttgvi-----ngad-----nmLldsgfyk 789
QY 502 TDYDYLQ---MSEPNKRYGDEDLRNALQAL 528
Db 790 tpkynlgnyywedtnkdqgkdstekglsigv 819

RESULT 25
AAW08192
ID AAW08192 standard; protein; 482 AA.
XX
XX
AC AAW08192;
XX
XX
DT 16-JUL-1997 (first entry)
DE
DE B. licheniformis mutant alpha-amylase, Asn 188 deletion.
XX
XX Asparagine; substitution; mutation; dish-washing; fabric-cleaning;

KW detergent.
XX Bacillus licheniformis.
OS Synthetic.
OS
PN W09639528-A2.
XX
PD 12-DEC-1996.
XX
PF 06-JUN-1996; 96WO-US09089.
XX
PR 06-JUN-1995; 95US-0468700.
XX
XX (GEMV) GENENCOR INT INC.
XX
PI Day A, Mitchinson C, Requaadt C, Ringer C, Ropp T;
PI Solheim LP;
XX
XX WPI; 1997-043150/04.
XX
PT Mutant alpha-amylase with improved low pH starch hydrolysis activity
PT - has at least one asparagine residue substituted or deleted, useful
PT in detergents
XX
PS Claim 2; Page -; 49pp; English.
XX
CC Novel mutant alpha-amylases contain a deletion or substitution of at
CC least one asparagine residue from the wild-type. The present sequence
CC is a specific example of such a mutant alpha-amylase from which the
CC Asn residue at position 188 has been deleted. The mutant alpha-amylase
CC is useful for starch processing and liquefaction. It is used in dish-
CC washing or fabric-cleaning detergent compositions. The modifications
CC produce enzymes with enhanced low pH performance suited for commercial
CC liquefaction conditions. The mutant alpha-amylases are also found to
CC have improved thermal and oxidative stability and to be more stable
CC in the absence of calcium ions compared to wild-type.
CC N.B. The present sequence is not shown in the specification but is
CC derived from the Bacillus licheniformis alpha-amylase sequence shown in
CC figure 4.
XX
SQ Sequence 482 AA;

Query Match 5.5%; Score 156.5; DB 18; Length 482;
Best Local Similarity 22.6%; Pred. No. 0.0058;
Matches 77; Conservative 40; Mismatches 119; Indels 105; Gaps 14;

QY 7 EFLIANDIDNSPVVQAEQNLNLYLLNFGTITANNDQANFDSVRVADPNIDADLMIAT 66
Db 193 dylmyadiydhpdvaeeikrw-----gtwyan--elqlgdfldav----- 232
QY 67 QDYFNAAYGMSDAYSNNKHII--NILEDW-NHADPEYFNKIGNPQLTMDTINKSL----- 118
Db 233 -----khikfsfirdwnhvr-----ektgkemtvaeywqndlgalen 271
QY 119 -----NHGLSDATNRWGLDAIVHQSLADRENNSSTENVIPNYSFVRADNNSQDQIQN 171
Db 272 ylnktnfnshsvifdphlyqfha-----astgggydmdckllngtvsakh----- 315
QY 172 AIRDVTGKDYHRTFEDEKGDVAYIQDQNSVVKYVNLVNIPIASVAILLTKNDITPRVY 231
Db 316 plksvtfvndh-----dtqpg-----qslstvcvfwk---playafiltresgypqvfy 362
QY 232 GDLYTGGQYMEHQTRYTYDTLTNLLKSRVKYVAGGQ-----SMQTMVSGVGNNN 279
Db 363 gdmlygtkgsdgreipalkhkiepilkartqaygaqdyfhdhdivgtvregdsavang 422
QY 280 ILTSVRYKGA-----MTATDTGTGTETRTQGTGGVYVWSN 312
Db 423 laalitdp9ggakrmvvgvgrnagetwhditgrnsepvvins 463

Search completed: August 12, 2002, 07:59:43
Job time: 260 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 12, 2002, 07:58:33 ; Search time 15.41 Seconds

(without alignments)
860.681 Million cell updates/sec

Title: US-09-995-749A-2_COPY_972_1514

Perfect score: 2852

Sequence: 1 QYRGGEFLANDIDNSNPV.....AIQALHKAGLQVMDRWPDQ 543

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : Issued Patents AA:*

1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep.*

2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep.*

3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep.*

4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep.*

5: /cgn2_6/ptodata/2/1aa/PTCUS.COMB.pep.*

6: /cgn2_6/ptodata/2/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1464.5	51.3	1375	4	US-09-210-361-4
2	1413	49.5	1430	3	US-09-088-172-2
3	1413	49.5	1430	4	US-09-210-361-6
4	1397	49.0	1475	3	US-09-007-999-2
5	1397	49.0	1475	4	US-09-210-361-2
6	1288.5	45.2	1577	2	US-08-793-824-2
7	165.5	5.8	750	3	US-08-814-052-2
8	165.5	5.8	750	3	US-08-812-829-2
9	158	5.5	560	3	US-08-814-052-6
10	158	5.5	560	3	US-08-812-829-6
11	157.5	5.5	485	4	US-09-291-023A-18
12	156.5	5.5	483	1	US-08-468-700-34
13	156.5	5.5	483	2	US-08-468-700-32
14	156.5	5.5	483	2	US-08-468-698-32
15	156.5	5.5	483	2	US-08-704-706A-34
16	156.5	5.5	483	2	US-08-600-908A-13
17	156.5	5.5	483	3	US-08-830-383-3
18	156.5	5.5	483	3	US-08-683-838A-13
19	156.5	5.5	483	3	US-08-914-679A-3
20	156.5	5.5	483	4	US-09-182-859-2
21	156.5	5.5	483	4	US-09-170-670-4
22	156.5	5.5	483	4	US-09-193-088-4
23	156.5	5.5	483	4	US-09-183-412-4
24	156.5	5.5	483	4	US-08-985-659-3s
25	156.5	5.5	483	4	US-09-264-097-2
26	156.5	5.5	483	4	US-08-194-664A-32
27	156.5	5.5	483	4	US-09-291-023A-21

28	156.5	5.5	483	5	PCT-US94-01553A-32	Sequence 32, Appl
29	156.5	5.5	483	5	PCT-US95-10426-32	Sequence 32, Appl
30	156.5	5.5	487	2	US-08-468-220-37	Sequence 37, Appl
31	156.5	5.5	487	2	US-08-468-698-37	Sequence 37, Appl
32	156.5	5.5	487	4	US-08-194-664A-37	Sequence 37, Appl
33	156.5	5.5	487	5	PCT-US94-01553A-37	Sequence 37, Appl
34	156.5	5.5	487	5	PCT-US95-10426-37	Sequence 37, Appl
35	156.5	5.5	511	1	US-08-468-700-35	Sequence 35, Appl
36	156.5	5.5	511	1	US-08-645-971-3	Sequence 3, Appl
37	156.5	5.5	511	2	US-08-468-220-33	Sequence 33, Appl
38	156.5	5.5	511	2	US-08-468-698-33	Sequence 33, Appl
39	156.5	5.5	511	2	US-08-704-706A-35	Sequence 35, Appl
40	156.5	5.5	511	3	US-08-890-383-2	Sequence 2, Appl
41	156.5	5.5	511	3	US-08-890-383-4	Sequence 4, Appl
42	156.5	5.5	511	3	US-08-914-679A-2	Sequence 2, Appl
43	156.5	5.5	511	3	US-08-914-679A-4	Sequence 4, Appl
44	156.5	5.5	511	4	US-08-985-659-36	Sequence 36, Appl
45	156.5	5.5	511	4	US-08-194-664A-33	Sequence 33, Appl
46	156.5	5.5	511	5	PCT-US94-01553A-33	Sequence 33, Appl
47	156.5	5.5	511	5	PCT-US95-10426-33	Sequence 33, Appl
48	156.5	5.5	512	1	US-07-623-953-3	Sequence 3, Appl
49	156.5	5.5	512	1	US-07-623-953-5	Sequence 5, Appl
50	156.5	5.5	512	1	US-08-720-899-2	Sequence 2, Appl
51	156.5	5.5	512	1	US-08-459-610-2	Sequence 2, Appl
52	156.5	5.5	512	2	US-08-343-804-2	Sequence 2, Appl
53	156.5	5.5	512	2	US-08-687-399-2	Sequence 2, Appl
54	156.5	5.5	512	2	US-08-600-908A-2	Sequence 2, Appl
55	156.5	5.5	512	3	US-08-683-838A-2	Sequence 2, Appl
56	156.5	5.5	512	4	US-08-985-659-34	Sequence 34, Appl
57	156.5	5.5	631	3	US-08-814-052-8	Sequence 8, Appl
58	156.5	5.5	631	3	US-08-812-829-8	Sequence 8, Appl
59	155.5	5.5	483	1	US-08-645-971-2	Sequence 2, Appl
60	154.5	5.4	484	6	5171673-8	Sequence 2, Appl
61	154.5	5.4	487	6	5171673-6	Sequence 2, Appl
62	153.5	5.4	483	2	US-08-468-220-36	Sequence 36, Appl
63	153.5	5.4	483	2	US-08-468-698-36	Sequence 36, Appl
64	153.5	5.4	483	5	US-08-194-664A-36	Sequence 36, Appl
65	153.5	5.4	483	5	PCT-US94-01553A-36	Sequence 36, Appl
66	153.5	5.4	483	5	PCT-US95-10426-36	Sequence 36, Appl
67	149.5	5.2	480	4	US-09-182-859-4	Sequence 4, Appl
68	149.5	5.2	480	4	US-09-170-670-5	Sequence 5, Appl
69	149.5	5.2	480	4	US-09-193-088-5	Sequence 5, Appl
70	149.5	5.2	480	4	US-09-183-412-5	Sequence 5, Appl
71	149.5	5.2	483	4	US-09-291-023A-16	Sequence 16, Appl
72	149.5	5.2	485	4	US-09-264-097-7	Sequence 7, Appl
73	149.5	5.2	514	1	US-08-720-899-4	Sequence 4, Appl
74	149.5	5.2	514	1	US-08-459-610-4	Sequence 4, Appl
75	149.5	5.2	514	2	US-08-343-804-4	Sequence 4, Appl
76	149.5	5.2	514	2	US-08-687-399-4	Sequence 4, Appl
77	149.5	5.2	514	2	US-08-600-908A-4	Sequence 4, Appl
78	149.5	5.2	514	3	US-08-683-838A-4	Sequence 4, Appl
79	149.5	5.2	514	4	US-09-264-097-4	Sequence 4, Appl
80	149.5	5.2	520	1	US-08-468-700-36	Sequence 36, Appl
81	149.5	5.2	520	1	US-08-645-971-4	Sequence 4, Appl
82	149.5	5.2	520	2	US-08-468-220-34	Sequence 34, Appl
83	149.5	5.2	520	2	US-08-468-698-34	Sequence 34, Appl
84	149.5	5.2	520	2	US-08-704-706A-36	Sequence 36, Appl
85	149.5	5.2	520	3	US-08-890-383-5	Sequence 5, Appl
86	149.5	5.2	520	3	US-08-914-679A-5	Sequence 5, Appl
87	149.5	5.2	520	4	US-08-985-659-37	Sequence 37, Appl
88	149.5	5.2	520	4	US-08-194-664A-34	Sequence 34, Appl
89	149.5	5.2	520	5	PCT-US94-01553A-34	Sequence 34, Appl
90	149.5	5.2	520	5	PCT-US95-10426-34	Sequence 34, Appl
91	146.5	5.1	485	2	US-08-446-803-1	Sequence 1, Appl
92	146.5	5.1	485	2	US-08-861-837-1	Sequence 1, Appl
93	146.5	5.1	485	2	US-08-600-908A-12	Sequence 12, Appl
94	146.5	5.1	485	3	US-08-683-838A-12	Sequence 12, Appl
95	146.5	5.1	485	3	US-08-600-908A-1	Sequence 1, Appl
96	146.5	5.1	485	4	US-09-170-670-1	Sequence 1, Appl
97	146.5	5.1	485	4	US-09-170-670-7	Sequence 7, Appl
98	146.5	5.1	485	4	US-09-193-088-1	Sequence 1, Appl
99	146.5	5.1	485	4	US-09-193-088-7	Sequence 7, Appl
100	146.5	5.1	485	4	US-09-183-412-1	Sequence 1, Appl

ALIGNMENTS

```

RESULT 1
US-09-210-361-4
; Sequence 4, Application US/09210361
; Patent No. 6284479
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starches and
; FILE REFERENCE: 0357CR
; CURRENT APPLICATION NUMBER: US/09/210,361
; CURRENT FILING DATE: 1998-12-11
; EARLIER APPLICATION NUMBER: 09/007,999
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/009,620
; EARLIER FILING DATE: 1998-01-20
; EARLIER APPLICATION NUMBER: 08/485,243
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/008,172
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 1375
; TYPE: PRT
; ORGANISM: streptococcus mutans
US-09-210-361-4

Query Match      51.38; Score 1464.5; Db 4; Length 1375;
Best Local Similarity 52.98; Pred. No. 1.4e-117; Indels 25; Gaps 8;
Matches 292; Conservative 94; Mismatches 141;

QY 1 QTYGQRELLANDIDNSNPVQAQQLNMLYLLNFGTITANNDQANFDSRVYDAPDNIDA 60
Db 425 RTIGGYEFLANDVNSNPVQAQQLNMLHFLMFGNIYANDPANEFSRVDAVDNVA 484
QY 61 DLMNIAQDYFNAAYGM-DSDAVSNKHINLEDDNHADPEYFNKIGNPOLMTDITKNSLA 119
Db 485 DLQIAGDYLAKAAGIHKNDKAANDHLSLEAWSYNTPTYLHDDGDNNIMNDKRLSLL 544
QY 120 HGLSDATN-RWGLDAIVHQSLADRENNSTENVIPNTSFVRAHDNNSQDQIQNAIR--- 174
Db 545 YSLAKPLAQRSQGNMPLNTSLNVRTDDNAETAAPVPSYFIRAHDSVQDLIRIIRTEIN 604
QY 175 -DVTGKDYHTTFDEQKIDAYIQDQNSTVKKYLNINIPASYAILTNKDTTPRVYIGD 233
Db 605 PRVVG---YSEFTEIKKAFELYNKDLATEKKYTHYNTALSALLTNKSSVPRVYIGD 661
QY 234 LYTDGQGYMEHQTRYDPTLTNLAKSRVKYVAGGSMQTMVSVGGNNILTSVRYKKGAMTA 293
Db 662 MFTDDGQYMAHTINYEALETLLKARIKYYVSGQAMRNQOV-GNSEIITSVRYKKGALKA 720
QY 294 TDTGDRTRTQIGVWVWNTPNLKVNDKVVHLHMGAAHKNQYRAAVLTDTGVIWNTS 353
Db 721 TDTGDRTRTQIGVWVWNTPNLKVNDKVVHLHMGAAHKNQYRAAVLTDTGVIWNTS 353
QY 354 DQAGAP--VAMTDENGDLVLSHNLVNGKEADTAVQGYANPDVSGYLAVWVPVAGSDNQ 411
Db 781 DQAGALVRYTNDRGELIFTA-----AD--IKGANPQVSGYGLVWVPVAGAADQ 829
QY 412 DARTAPSTKNSGAYRTNAAFDSNVIFEAFSNFYVPTTKESERANVRIAAQNADFASL 471
Db 829 DVRVAASTPSTDGKSVHQNALDSRVMEFESNFQAFATKKEEYTNVVIKNDYKFAEW 888

```

```

QY 472 GFTSFEMAPQYNSSKDRFTFLDSTIDNGYAFTRDYDLGMSFPNKYGTDEDLNAIQALHKA 531
Db 889 GVTDFEMAPQYVSSDTGSLDSVLQNGYAFTRDYDLGMSFPNKYGTADDLVKAIKALHRSK 948
QY 532 GLOWNADWVPDQ 543
Db 949 GIKVNADWVPDQ 960

```

```

RESULT 2
US-09-008-172-2
; Sequence 2, Application US/09008172
; Patent No. 6127802
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starch and
; FILE REFERENCE: 0358D
; CURRENT APPLICATION NUMBER: US/09/008,172
; CURRENT FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1430
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-008-172-2

```

```

Query Match      49.58; Score 1413; Db 3; Length 1430;
Best Local Similarity 50.38; Pred. No. 4.2e-113; Indels 34; Gaps 8;
Matches 282; Conservative 96; Mismatches 149;

QY 5 GOEFLANDIDNSNPVQAQQLNMLYLLNFGTITANNDQANFDSRVYDAPDNIDA 64
Db 417 GYDFELLANDIDNSNPVQAQQLNMLYLLNFGTITANNDQANFDSRVYDAPDNIDA 476
QY 65 IAQDYFNAAYGM-SDAVSNKHINLEDDNHADPEYFNKIGNPOLMTDITKNSLA 123
Db 477 IASDYLAHGYDKSEKNAHNLHLSLEAWSNDPOYKDTKGALPDLNKLRLSLLALT 536
QY 124 -----DAVIN---RWGLDAIVHQSLADRENNSTENVIPNTSFVRAHDNNSQDQIQNAIR 174
Db 537 RPLDASAKNKEIRSLPEVITNSLNRSAGEKNSERMANIFIRAHDSVQTVIAKLIK 596
QY 175 -DVTGK-DYHTTFDEQKIDAYIQDQNSTVKKYLNINIPASYAILTNKDTTPRVYIG 232
Db 597 RQINPKTQGLTFTLDELKQAFKYYNEDKQAKKYYTQSNIPITAYALMLSNKDSITRLYYG 656
QY 233 DLYTDGQGYMEHQTRYDPTLTNLAKSRVKYVAGGSMQTMVSVGGNNN-----ILTSV 284
Db 657 DMYSDGQYMATKSPYYDAIDTLLKARIKAAAGQDMKITVEGDKSHMDWDYTGVLTSV 716
QY 285 RYKKGAMTADGTGTRFQIGVWVWNTPNLKVNDKVVHLHMGAAHKNQYRAAVLT 344
Db 717 RYKKGAMTADGTGTRFQIGVWVWNTPNLKVNDKVVHLHMGAAHKNQYRAAVLT 344
QY 345 TQGVINVTSDQAGAP--VAMTDENGDLVLSHNLVNGKEADTAVQGYANPDVSGYLAVW 402
Db 777 KGLTFTTSDAAAKSLYKTNKNGELVFDASD-----IQGLNLPQVSGYLAVW 824
QY 403 VPVAGSDNQDARTAPSTKNSGAYRTNAAFDSNVIFEAFSNFYVPTTKESERANVRI 462
Db 825 VPVAGSDNQDVRVAASNRKANATGQYVESSALDSQLIYEGFSNFQDFVTKDSYTNKKA 884
QY 463 QNADFASLGTFSFEMAPQYNSSKDRFTFLDSTIDNGYAFTRDYDLGMSFPNKYGTDEDL 522
Db 885 QNADFASLGTFSFEMAPQYNSSKDRFTFLDSTIDNGYAFTRDYDLGMSFPNKYGTDEDL 522
QY 523 NAIQALHKAAGLOWNADWVPDQ 543

```



```
QY 98 GYVLKADGGYYNLTGTTTQKFLPIQLIGKKQNGFVKGNDGNYFYDLGNVNTFI 157
|||
Db 184 GYVYVSDGSY-----SKDFEKIN-GTWYFDGSGYMLSDRWK 221
|||
QY 158 EDVSGNMYFDDQD-----KVENKHFDVDSYGEKG-----YFFI--KNGVSFRG 202
|||
Db 222 KHTDGNWYFDGGEWATGWKKADKXKWFYFDVGAMKTGWVKYKDTWYLDAREGAMVSN 281
|||
QY 203 GLVOTDNGT--YYFDNFGKVMRNOTINAGAMITLDENG 239
|||
Db 282 AFIQSADGTGWYLLKPDGTL-----ADKPEFTVDPDG 313
|||
RESULT 8
ALYS_STRPN
ID Alys_STRPN STANDARD; PRT; 318 AA.
AC P06653;
DT 01-JAN-1988 (Rel. 06, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Autolysin (EC 3.5.1.28) (N-acetylmuramoyl-L-alanine amidase) (Murein
hydrolyase) (Mucopolysaccharide aminohydrolase) (Cell wall hydrolase).
GN LYTA OR SP1937.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86301881; PubMed=2875013;
RA Garcia P., Garcia J.L., Garcia E., Lopez R.;
RT "Nucleotide sequence and expression of the pneumococcal autolysin
gene from its own promoter in Escherichia coli.";
RL Gene 43:265-272(1986).
[2]
RP SEQUENCE FROM N.A.
RX STRAIN=R800;
RA Martin B.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
[3]
RP SEQUENCE FROM N.A.
RX STRAIN=TIGR4;
RC MEDLINE=21357209; PubMed=11463916;
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
Peterson S., Heidelberg J., Deboy R.T., Haft D.H., Dodson R.J.,
Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
Holtzapple E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,
McDonald L.A., Feldblyum T.V., Anguoli S., Dickinson T., Hickey E.K.,
Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C., C.M.;
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
pneumoniae.";
RL Science 293:498-506(2001).
CC -1- FUNCTION: AUTOLYSINS ARE INVOLVED IN SOME IMPORTANT BIOLOGICAL
PROCESSES SUCH AS CELL SEPARATION, CELL-WALL TURNOVER, COMPETENCE
FOR GENETIC TRANSFORMATION, FORMATION OF THE FLAGELLA AND
SPORULATION. AUTOLYSIN STRICTLY DEPEND ON THE PRESENCE OF CHOLINE-
CONTAINING CELL WALLS FOR ACTIVITY.
CC -1- CATALYTIC ACTIVITY: Hydrolyzes the link between N-acetylmuramoyl
residues and L-amino acid residues in certain bacterial cell-wall
glycopeptides.
CC -1- MISCELLANEOUS: THE C-TERMINAL DOMAIN COULD BE RESPONSIBLE FOR THE
SUBSTRATE RECOGNITION.
CC -1- SIMILARITY: BELONGS TO THE N-ACETYLMURAMOYL-L-ALANINE AMIDASE
FAMILY 2.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
```

```
or send an email to license@isb-sib.ch).
CC EMBL; M13812; AAR26917.1; -
DR EMBL; Z34303; CA844074.1; -
DR EMBL; AF007483; AAK76005.1; -
DR PIR; A25634; A25634.
DR TIGR; SP1937.
DR InterPro; IPR002502; Amidase_2.
DR InterPro; IPR002479; CW_binding.
DR Pfam; PF01510; Amidase_2; 1.
DR Pfam; PF01473; CW_binding_1; 4.
KW Hydrolase; Cell wall; Repeat; Complete proteome.
FT DOMAIN 177 302 6 X 20 AA TANDEM REPEATS.
FT REPEAT 177 196 1.
FT REPEAT 197 218 2.
FT REPEAT 219 239 3.
FT REPEAT 240 259 4.
FT REPEAT 260 282 5.
FT REPEAT 283 302 6.
FT CONFLICT 304 304 K -> R (IN REF. 1 AND 2).
SQ SEQUENCE 318 AA; 36544 MW; 3CA7F3CD132FB502 CRC64;

Query Match 9.2%; Score 132; DB 1; Length 318;
Best Local Similarity 21.9%; Pred. No. 0.003;
Matches 62; Conservative 44; Mismatches 99; Indels 78; Gaps 14;

QY 8 EVATVTRVDRGNVWDAIINNLYVNTTIGGGEYQKKGGA----- 49
|||
Db 48 ELGFFSHVGVGCMQVGPVDNGAW---DVGGWNAETAAVELIESHSTKEEFTMDYRL 104
|||
QY 50 FLDKLQKLYPEIFTKKOVSTGVA-----IDPSQKITEN--SAKYF 87
|||
Db 105 YIELLRNLADAGLPKTLDTGSLAGIKTHEYCTNNQPNNSHDVDPYLLAKWGISHEQF 164
|||
QY 88 -----NCTNLIHRSGYVLKADGGYYNLTGTTTQKFLPIQLIGKKQNGFVKGNDGNY 142
|||
Db 165 KHTIENGLTI---ETGW-QKNDTGWY-----VHSDGSPKDKFEKIN-GTW 206
|||
QY 143 YFYDLAGNMVKNFTIEDSVGNWYFFDQDGKVEN-KHFVDVDSYGEKGYFFFLKNGVSFR 201
|||
Db 207 YFYDSSGYMLADRWKHTDGNWYFDNDSGEMATGWKKIAD-----KWYFFNEEG-AMK 258
|||
QY 202 GGLVOTDNGTYIFD-NYGMKVMRNOTINA--GMIYITLDENGKL 241
|||
Db 259 TGWYKXKDTWYLDAREGAMVSNAFIQSADGTGWYLLKPDGTL 301
|||
RESULT 9
LYTB_STRPN
ID LYTB_STRPN STANDARD; PRT; 658 AA.
AC Q924P7;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Putative endo-beta-N-acetylglucosaminidase precursor (EC 3.2.1.96)
(Murein hydrolase).
GN LYTB OR SP0965.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 24-30.
RC STRAIN=R6;
RX MEDLINE=99195827; PubMed=10096093;
RA Garcia P., Gonzalez M.P., Garcia E., Lopez R., Garcia J.L.;
RT "LytB, a novel pneumococcal murein hydrolase essential for cell
separation.";
RL Mol. Microbiol. 31:1275-1281(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=TIGR4;
```


RX MEDLINE-21357209; PubMed-114633916;
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
RA Urquay A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Dunham L.A., White O., Salzberg S.L., Lewis M.R., Redune D.,
RA Holtzapple E., Khouri H., Wolf A.M., Uterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum F.V., Angluvo S., Dickinson V., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae";
RL Science 293:498-506(2001).
CC -|- FUNCTION: PLAYS AN IMPORTANT ROLE IN CELL WALL DEGRADATION AND
CC CELL SEPARATION.
CC -|- CATALYTIC ACTIVITY: Endohydrolysis of the di-N-acetylchitobiosyl
CC unit in high-mannose glycopeptides and glycoproteins containing
CC the -[Man(GlcNAc)2]Asn-structure. One N-acetyl-D-glucosamine
CC residue remains attached to the protein; the rest of the
CC oligosaccharide is released intact.
CC -|- SIMILARITY: BELONGS TO FAMILY 73 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ010312; CA009078.1; -;
DR EMBL; AE007400; ANK75086.1; -;
DR TIGR; SF0965; -;
DR InterPro; IPR002901; Amidase_4.
DR Pfam; PF01832; Amidase_4; 1.
DR SMART; SM00047; LY2; 1.
KW Signal; Hydrolase; Cell wall; Complete proteome.
FT SIGNAL 1 23
FT CHAIN 24 658 PUTATIVE ENDO-BETA-N-
FT ACETYLGLUCOSAMINIDASE.
FT CONFLICT 336 336 I -> M (IN REF. 1).
FT CONFLICT 381 381 T -> A (IN REF. 1).
FT CONFLICT 384 384 E -> K (IN REF. 1).
FT CONFLICT 535 535 L -> P (IN REF. 1).
FT CONFLICT 580 580 F -> S (IN REF. 1).
SQ SEQUENCE 658 AA; 76469 MW; B625515006C9CB36 CRC64;

Query Match 9.0%; Score 128.5; DB 1; Length 658;
Best Local Similarity 25.5%; Pred. No. 0.013;
Matches 50; Conservative 25; Mismatches 64; Indels 57; Gaps 10;
QY 74 DPSQKITWSAKYFNGNTHLRSGYVLKADGGQYNNLGTGTPQLTGEKKQGNQEG 133
DB 37 DQSQANNEW---VFD---THYQSMFYIKADANY-----AENEW 68
QY 134 FVKGNDGNYFYDLAGNVMKNFTFEDSGNVMKFNKHPVDVDSYGEKG----- 189
DB 69 LKQDD--YFLKSGGYWAKSEWEDK-GARYLDQDGMKRNNA-WWTSTYGTAGAKVI 124
QY 190 -----YVFFLK-NGVSFRGLGVTDNGTYTFDYGKKNRQTIN-----AGAMI 232
DB 125 EDWVYDSQYDAWFYIKADQSHAENELIKGDKDYFKSGGYLLLSQWLNQAVVNASGAKV 184
QY 233 YTLDENGKLIKASYSN 248
DB 185 -----QQGMLEDKQYQS 196
RESULT 10
TOXA_CLODI
ID TOXA_CLODI STANDARD; PRT; 2710 AA.
AC P16154;
DT 01-APR-1990 (Rel. 14, Created)

DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Toxin A.
GN TOXA OR TCDA.
OS Clostridium difficile.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxId=1496;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-VPI 10463;
RX MEDLINE-90221894; PubMed-2109310;
RA Sauerborn M., von Eichel-Streiber C.;
RT "Nucleotide sequence of Clostridium difficile toxin A";
RL Nucleic Acids Res. 18:1629-1630(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-VPI 10463;
RX MEDLINE-90129305; PubMed-2105276;
RA Dove C.H., Wang S.Z., Price S.B., Phelps C.J., Iyerly D.M.,
RA Wilkins T.W., Johnson J.L.;
RT "Molecular characterization of the Clostridium difficile toxin A
RT gene";
RL Infect. Immun. 58:480-488(1990).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-VPI 10463;
RA Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
CC -|- DOMAIN: THE C-TERMINAL PART OF TOXIN A CONSISTS OF A 833 AA
CC REPETITIVE STRUCTURE. THIS PART OF TOXIN A IS COMPOSED OF FIVE
CC DIFFERENT OLIGOPEPTIDES.
CC -|- DISEASE: CLOSTRIDIUM DIFFICILE PRODUCES TWO TOXINS, AN
CC ENTEROTOXIN CALLED A AND CYTOTOXIN B. ONLY AFTER THE ENTERAL
CC DELIVERY OF THE ENTEROTOXIN A MAY THE CHARACTERISTIC DISEASE
CC CALLED PSEUDOMEMBRANOUS COLITIS BE INDUCED.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X51797; CAA36094.1; -;
DR EMBL; M30307; AAA23283.1; -;
DR EMBL; X92982; CAA63564.1; -;
DR PIR; S08638; S08638.
DR InterPro; IPR002479; CW_binding.
DR Pfam; PF01473; CW_binding_1; 31.
KW Toxin; Enterotoxin.
SQ SEQUENCE 2710 AA; 308052 MW; 0A6E52CE84C14421 CRC64;

Query Match 8.7%; Score 125; DB 1; Length 2710;
Best Local Similarity 24.1%; Pred. No. 0.14;
Matches 68; Conservative 37; Mismatches 101; Indels 76; Gaps 17;
QY 16 DDGNVWRKDAI-INNNLYVNV-----TIGGGEYQKKYGGAFDLKQLYPEIFT 63
DB 1824 DEDSKLVKGLININSLFYFDPISFNLVGTGQTNG---KKY---YFD----- 1865
QY 64 KKQVSTGVADPSQKITWSAKYFNGNTHLRSGYVLKADGGQYNNLGTGTPQLP-- 120
DB 1866 ---INTGAAL-TSYKIIKGKHFYFNNDGVNQLG---YFKGPDGDFEYFAPANTQNNIEGQ 1918
QY 121 -----IQLTGEKK--QGNQEGYVKG-----DGNVYFYDLAGNVMKNFTFEDSVGNWYF 166
DB 1919 AIVYQSKFTLNGKKYFDNNSKAVTGWRIINNEKIYFN--PNNIAVGLQVIDNNKY 1976
QY 167 FDQDCKMWE-----NKHFDVDVDS-----YGEK---GTFFFLKNGVSFRGLVQTONGT 211
DB 167 FDQDCKMWE-----NKHFDVDVDS-----YGEK---GTFFFLKNGVSFRGLVQTONGT 211

Dd 1977 FNPDPALISGHWGTVNGSRYYFTDPAIAFANGKTKTDGRHFFYDSOCVKIGVFSTSNCF 2036

Oy 212 YIF--DNYGKMVRNQTINAGAMIYTIDENGKLIRASYNSDA 250
II : I : I : I : I : II III

Dd 2037 EYEAPANTNNTEGAIVQSKFLTL--NGKKYPDNNSKA 2076

RESULT 11

ID WAPA_BACSU STANDARD; PRT; 2334 AA.

AC Q07833;

DT 01-OCT-1994 (Rel. 30, Created)

DT 01-OCT-1994 (Rel. 30, Last sequence update)

DT 01-MAR-2002 (Rel. 41, Last annotation update)

DE Wall-associated protein precursor.

GN WAPA OR NIYG.

OS Bacillus subtilis.

CC Bacteria; Firmicutes; Bacillus/Clostridium group;
Bacillus/Staphylococcus group; Bacillus.
NCBI_TaxId=1423;

RX [1]

RP SEQUENCE FROM N.A.

RC STRAIN=168;

RX MEDLINE=93302506; PubMed=8316082;

RA Foster S.J.;

RT "Molecular analysis of three major wall-associated proteins of
Bacillus subtilis 168: evidence for processing of the product of a
gene encoding a 258 kDa precursor two-domain ligand-binding
protein.";

KT Mol. Microbiol. 8:299-310(1993).

RL [2]

RN SEQUENCE FROM N.A.

RP STRAIN=168 / BGSC1al;

RC MEDLINE=95219088; PubMed=7704263;

RX Yoshida K.-I., Sano H., Seki S., Fujimura M., Fujita Y.;
"Cloning and sequencing of a 29 kb region of the Bacillus subtilis
genome containing the hut and wapa loci.";

RA Microbiology 141:337-343(1995).
[3]

RN SEQUENCE FROM N.A.

RP STRAIN=168 / BGSC1al;

RC MEDLINE=97124196; PubMed=8969509;

RX Yoshida K.-I., Shindo K., Sano H., Seki S., Fujimura M., Yanai N.,
Miwa Y., Fujita Y.;
"Sequencing of a 65 kb region of the Bacillus subtilis genome
containing the lic and cel loci, and creation of a 177 kb contig
covering the gnt-sacxy region.";

KT Microbiology 142:3113-3123(1996).

RL [1-]
FUNCTION: STILL UNKNOWN NOT INVOLVED IN CELL MEMBRANE METABOLISM,
MOTILITY, SECRETION OR DIFFERENTIATION.

CC -1 SUBCELLULAR LOCATION: MEMBRANE-ASSOCIATED. MAY BE RELEASED
INTO THE MEDIUM.

CC -1 DOMAIN: HAS TWO LIGAND-BINDING DOMAINS; THE N-TERMINUS, HAS THREE
101 AA REPEATS WHICH ARE RESPONSIBLE FOR CELL WALL BINDING; THE
C-TERMINUS CONSISTS OF TWO BLOCKS OF RESIDUES WITH A CONSERVED
MOTIF REPEATED 31 TIMES.

CC -1 SIMILARITY: THE REPEATED MOTIF IN THE C-TERMINUS HAS SOME
SIMILARITY TO THE REPEAT IN E.COLI RHS GROUP OF PROTEINS (RHS-A-D).

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to licensel@lsb-sib.ch).

DR ENBL; L05634; AAA22883.1; *

DR ENBL; D31856; BAA06656.1; *

DR ENBL; D23985; BAA06260.1; *

DR ENBL; D84026; BAA11683.1; *

DR ENBL; Z99124; CAB15959.1; *

DR PIR; S32920; S32920.

DR	SubtilList; BG10797; wapa.	
DR	InterPro; IPR003305; CBD_6.	
DR	Pfam; PF02018; CBD_6; 1.	
KW	Cell wall; Repeat; Signal;	
FT	SIGNAL	3
FT	CHAIN	29
FT	DOMAIN	29
FT	DOMAIN	504
FT	REPEAT	504
FT	REPEAT	636
FT	REPEAT	769
FT	DOMAIN	1021
FT	REPEAT	1021
FT	REPEAT	1040
FT	REPEAT	1061
FT	REPEAT	1063
FT	REPEAT	1083
FT	REPEAT	1109
FT	REPEAT	1129
FT	REPEAT	1150
FT	REPEAT	1150
FT	REPEAT	1174
FT	REPEAT	1199
FT	REPEAT	1219
FT	REPEAT	1646
FT	REPEAT	1657
FT	REPEAT	1690
FT	REPEAT	1711
FT	REPEAT	1732
FT	REPEAT	1753
FT	REPEAT	1795
FT	REPEAT	1820
FT	REPEAT	1840
FT	REPEAT	1861
FT	REPEAT	1887
FT	REPEAT	1906
FT	REPEAT	1929
FT	REPEAT	1969
FT	REPEAT	1983
FT	REPEAT	2008
FT	REPEAT	2028
FT	REPEAT	2051
FT	REPEAT	2071
FT	REPEAT	2093
FT	REPEAT	2120
FT	SEQUENCE	2139
DR	SubtilList; BG10797; wapa.	
DR	InterPro; IPR003305; CBD_6.	
DR	Pfam; PF02018; CBD_6; 1.	
KW	Cell wall; Repeat; Signal;	
FT	SIGNAL	3
FT	CHAIN	29
FT	DOMAIN	29
FT	DOMAIN	504
FT	REPEAT	504
FT	REPEAT	636
FT	REPEAT	769
FT	DOMAIN	1021
FT	REPEAT	1021
FT	REPEAT	1040
FT	REPEAT	1061
FT	REPEAT	1063
FT	REPEAT	1083
FT	REPEAT	1109
FT	REPEAT	1129
FT	REPEAT	1150
FT	REPEAT	1150
FT	REPEAT	1174
FT	REPEAT	1199
FT	REPEAT	1219
FT	REPEAT	1646
FT	REPEAT	1657
FT	REPEAT	1690
FT	REPEAT	1711
FT	REPEAT	1732
FT	REPEAT	1753
FT	REPEAT	1795
FT	REPEAT	1820
FT	REPEAT	1840
FT	REPEAT	1861
FT	REPEAT	1887
FT	REPEAT	1906
FT	REPEAT	1929
FT	REPEAT	1969
FT	REPEAT	1983
FT	REPEAT	2008
FT	REPEAT	2028
FT	REPEAT	2051
FT	REPEAT	2071
FT	REPEAT	2093
FT	REPEAT	2120
FT	SEQUENCE	2139
DR	SubtilList; BG10797; wapa.	
DR	InterPro; IPR003305; CBD_6.	
DR	Pfam; PF02018; CBD_6; 1.	
KW	Cell wall; Repeat; Signal;	
FT	SIGNAL	3
FT	CHAIN	29
FT	DOMAIN	29
FT	DOMAIN	504
FT	REPEAT	504
FT	REPEAT	636
FT	REPEAT	769
FT	DOMAIN	1021
FT	REPEAT	1021
FT	REPEAT	1040
FT	REPEAT	1061
FT	REPEAT	1063
FT	REPEAT	1083
FT	REPEAT	1109
FT	REPEAT	1129
FT	REPEAT	1150
FT	REPEAT	1150
FT	REPEAT	1174
FT	REPEAT	1199
FT	REPEAT	1219
FT	REPEAT	1646
FT	REPEAT	1657
FT	REPEAT	1690
FT	REPEAT	1711
FT	REPEAT	1732
FT	REPEAT	1753
FT	REPEAT	1795
FT	REPEAT	1820
FT	REPEAT	1840
FT	REPEAT	1861
FT	REPEAT	1887
FT	REPEAT	1906
FT	REPEAT	1929
FT	REPEAT	1969
FT	REPEAT	1983
FT	REPEAT	2008
FT	REPEAT	2028
FT	REPEAT	2051
FT	REPEAT	2071
FT	REPEAT	2093
FT	REPEAT	2120
FT	SEQUENCE	2139
DR	SubtilList; BG10797; wapa.	
DR	InterPro; IPR003305; CBD_6.	
DR	Pfam; PF02018; CBD_6; 1.	
KW	Cell wall; Repeat; Signal;	
FT	SIGNAL	3
FT	CHAIN	29
FT	DOMAIN	29
FT	DOMAIN	504
FT	REPEAT	504
FT	REPEAT	636
FT	REPEAT	769
FT	DOMAIN	1021
FT	REPEAT	102

```

Query Match      8.6%; Score 123; DB 1; Length 2334;
Best Local Similarity 23.9%; Pred. No. 0.16;
Matches 68; Conservative 33; Mismatches 76; Indels 108; Gaps 17;

QY   28 NNNLYVVNTIGGGEVOKKYGCAFIDKKLKL-----YPFIFFKKQYSTGVAI 73
    || :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db   1901 NGNETTINSSASGRKVQVEYG-----KLNQLVKETHEDGTVEIYFDGFGNKRKTVT-- 1951
                                     :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
QY   74 DPSQKITWSAKYFNGT-NILHR-----GSGYVLKADGGQYY-----NLG 112
    | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db   1952 -----IKDGSCKTYNASFINNQLTKYNDESISYDKNGN---RFSOCKFTYTWDADENLT 2003
                                     :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
QY   113 TTTKFOLPIQLTGEBKKOGNEGFVRKNGNY--YFYDLAGNVAKVNFTIEDSYGNMYFFDDQ 170
    || :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db   2004 AVTK-----KGEDKPFATYKYDEKGNRIQKT-VNGKVTNFYF--D 2040
                                     :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
QY   171 GKMYENKHFEVD-----SYGSKCYFFFLKNGVSFRGLGVQTDNGTYTFPDNGKMWVRN 223
    | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db   2041 GDSLNVIYETDADNNVTSKYTYGSDGQL-----LSYTENGKKPYFHY----- 2082
                                     :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
QY   224 QTINA-GAMIVTLDBENGLIKASYNSDA-EYPTSTDVGKMLDQNK 266
    || :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db   2083 ---NAHGDIHIAISDSTGKTV-AKIQYDWANGAPTTEASDEVKONR 2123
                                     :|| :|| :|| :|| :|| :|| :|| :|| :|| :||

RESULT 12
TOXB CLODI
```

TOXIN_CLODI STANDARD; PRT: 2366 AA.
 AC P18177;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE Toxin B.
 GN TOXB OR TCDB.
 OS Clostridium difficile.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
 OC Clostridium.
 OX NCBI_TaxID=1496;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=VPI 10463;
 RX MEDLINE=90326540; PubMed=2374729;
 RA Barroso L.A., Wang S.Z., Phelps C.J., Johnson J.L., Wilkins T.D.;
 RT "Nucleotide sequence of Clostridium difficile toxin B gene."
 RL Nucleic Acids Res. 18:4004-4004(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=VPI 10463;
 RX MEDLINE=92293124; PubMed=1603068;
 RA Eichel-Streiber C., Lautenberg-Feidmann R., Sartingen S.,
 RA Schulze J., Sauerborn M.;
 RT "Comparative sequence analysis of the Clostridium difficile toxins A
 and B."
 RL Mol. Gen. Genet. 233:260-268(1992).
 CC -1- DISEASE: CLOSTRIDIUM DIFFICILE PRODUCES TWO TOXINS, AN
 ENTEROTOXIN CALLED A AND CYTOTOXIN B.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL: X53138; CAA37298.1;
 CC EMBL: X92982; CAA63562.1;
 CC EMBL: X60984; CAA43299.1;
 CC PIR: S10317; S10317.
 CC InterPro: IPR002479; CW_binding.
 CC Pfam: PF01473; CW_binding_1; 18.
 CC Cytotoxin; Toxin.
 CC SEQUENCE 2366 AA; 269709 MW; E1024BD8B8A56ADF CRC64;
 CC
 CC Query Match 8.1%; Score 115.5; DB 1; Length 2366;
 CC Best Local Similarity 24.5%; Pred. No. 0.62;
 CC Matches 68; Conservative 39; Mismatches 102; Indels 69; Gaps 15;
 CC
 CC 27 INNLYVNTIGGEGYQKYGAFLOKLYPEITTKQV-STGVAIDPSQIKTWSAK 85
 CC 1859 VGDDKYFNPINGG--AASTGETIID--DKNY--YFNQSGVLTQGVF-----STEDGFK 1906
 CC
 CC 86 YENCTNLL-----HRGSGYVLKADGGYYNLGTTTKQF 118
 CC 1907 YFAPANFDENLEGEAIDFTGKLIIDENIYFDNTYKGAIVENKELDGEHNYFSPKGFAP 1966
 CC
 CC 119 LPIQLTGKTKQ-----GNGEFVGKNGDGNYYFYDLAGNNVKNFTFSDSYGNWYFFDQDG 171
 CC 1967 KGLNQIGDYKYFNSDGMVGKGFVSINDKNHFDGDSGMVKGVTETD---GKHFFYAENG 2023
 CC
 CC 172 KM---VEN-----KHFV--DVDSYGEKGTFFELKNGVSFRGGLVQDNGTGYYP-DNKGKM 220
 CC 2024 EMQIGVNTEDGKFFAHHNEDLGNBS-----EELSY-SGILNFNNKIYFDOSGTAV 2076

QY 221 VRNQTINAGMIYILDENGK-LIKASVNSDAEYPTSTD 257
 DB 2077 VGWKDLEDGSKYFEDTAEAYIGLSLMDQGYFEND 2114
 RESULT 13
 LYCA_BPCP9 STANDARD; PRT: 339 AA.
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Lysozyme (EC 3.2.1.17) (Endolysin) (Muremidase) (CP-9 lysin).
 GN CPl9.
 OS Bacteriophage Cp-9.
 OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Podoviridae.
 OX NCBI_TaxID=10749;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90185251; PubMed=2311937;
 RA Garcia P., Garcia J.L., Garcia E., Sanchez-Puelles J.M., Lopez R.;
 RT "Modular organization of the lytic enzymes of Streptococcus
 pneumoniae and its bacteriophages."
 RL Gene 86:81-88(1990)
 CC -1- FUNCTION: RESPONSIBLE FOR THE SEPARATION OF THE HOST DAUGHTER
 CC CELLS AT THE END OF CELL DIVISION AND PARTICIPATES IN THE
 CC LIBERATION OF PROGENY BACTERIOPHAGE INTO THE MEDIUM. CPL-9
 CC STRICTLY DEPEND ON THE PRESENCE OF CHOLINE-CONTAINING CELL
 CC WALLS FOR ACTIVITY.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of the 1,4-beta-linkages between N-
 CC acetyl-D-glucosamine and N-acetylmuramic acid in peptidoglycan
 CC heteropolymers of the prokaryotes cell walls.
 CC -1- MISCELLANEOUS: THE C-TERMINAL DOMAIN COULD BE RESPONSIBLE FOR THE
 CC SUBSTRATE RECOGNITION.
 CC -1- SIMILARITY: BELONGS TO FAMILY 25 OF GLYCOSYL HYDROLASES.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL: M34760; AAA72845.1;
 CC PIR: JQ0438; MUBPC9.
 CC InterPro: IPR002479; CW_binding.
 CC InterPro: IPR002053; Glycosyl_hydrolase_25.
 CC Pfam: PF01473; CW_binding_1; 4.
 CC Pfam: PF01183; Glyco_hydro_25; 1.
 CC ProDom: PD004620; Glycosyl_hydrolase_25; 1.
 CC PROSITE: PS00953; GLYCOSYL_HYDROL_F25; 1.
 CC Hydrolase; Glycosidase; Bacteriolytic enzyme; Repeat.
 CC ACT_SITE 10 10
 CC ACT_SITE 37 37 BY SIMILARITY.
 CC ACT_SITE 6 X 20 AA TANDEM REPEATS.
 CC DONAIN 201 324
 CC REPEAT 201 221
 CC REPEAT 222 242
 CC REPEAT 243 262
 CC REPEAT 263 282
 CC REPEAT 283 304
 CC REPEAT 305 324
 CC SEQUENCE 339 AA; 39150 MW; A2143F05BDE4EDD CRC64;
 CC
 CC Query Match 8.0%; Score 114.5; DB 1; Length 339;
 CC Best Local Similarity 22.8%; Pred. No. 0.071;
 CC Matches 52; Conservative 44; Mismatches 85; Indels 47; Gaps 11;
 CC
 CC 58 YBEIFTKKQVSTGVAIDPSQIKTE-----WSAKYFNGTNILHRGSGYVLKADGGY--YN 110
 CC 120 YTPPIYSYKPFITLDNDYDQQLAQFPNLSLWAGY--GLNDGNADFEYFSPMCIIRWQYS 177

QY 111 LGTTTKQFLPI-----QLTGEKKQGNF-FVKGNQGN-----YFYDLA 149
 Db 178 SNPFDRNVLDDDEDEKPKTAGTWKQSKGWHFRNNGSFYKWKKEIGGVWYFDSKG 237
 QY 150 NMVKNFTFIEDSVGNWYFFQDQDKMVENKHFVDVDSYGERGTFFFLKNGVSFRGLVQTDN 209
 Db 238 YCITSEWLKDN-EKWTYLDKNGAMVTG--WVLVGS-----EWYMDSDGAMVTGWVKYKN 289
 QY 210 GTTYFND-YGKVRNQTINAGMIYTLDECKLIKASYNDAEYPTST 256
 Db 290 NWYMTNERNGNMVSNEFKSGKGYFMTNGEL-----ADNPSFT 329

RESULT 14

LYCA_BPCPI STANDARD; PRT: 339 AA.
 AC P15057; Q38009;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-MAR-2002 (Rel. 41, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Lysozyme (EC 3.2.1.17) (endolysin) (Muramidase) (CP-1 lysin).
 GN CPl1 OR 22.
 OS Bacteriophage Cp-1.
 OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Podoviridae.
 OX NCBI_TaxID=10747;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88124951; PubMed=3422470;
 RA Garcia E., Garcia J.L., Garcia P., Arraras A., Sanchez-Puelles J.M., Lopez R.;
 RT "Molecular evolution of lytic enzymes of Streptococcus pneumoniae and its bacteriophages";
 RL Proc. Natl. Acad. Sci. U.S.A. 85:914-918(1988).
 RN [2]

RP SEQUENCE FROM N.A.
 RX MEDLINE=96211500; PubMed=8648702;
 RA Martin A.C., Lopez R., Garcia P.;
 RT "Analysis of the complete nucleotide sequence and functional organization of the genome of Streptococcus pneumoniae bacteriophage Cp-1";
 RL J. Virol. 70:3678-3687(1996).
 CC -!- FUNCTION: RESPONSIBLE FOR THE SEPARATION OF THE HOST DAUGHTER CELLS AT THE END OF CELL DIVISION AND PARTICIPATES IN THE LIBERATION OF PROGENY BACTERIOPHAGE INTO THE MEDIUM. CPl-1 STRICTLY DEPEND ON THE PRESENCE OF CHOLINE-CONTAINING CELL WALLS FOR ACTIVITY.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of the 1,4-beta-linkages between N-acetyl-D-glucosamine and N-acetylmuramic acid in peptidoglycan heteropolymers of the prokaryotes cell walls.
 CC -!- MISCELLANEOUS: THE C-TERMINAL DOMAIN COULD BE RESPONSIBLE FOR THE SUBSTRATE RECOGNITION.
 CC -!- SIMILARITY: BELONGS TO FAMILY 25 OF GLYCOSYL HYDROLASES.

This SWISS-PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: J03586; AAA32204.1; -
 EMBL: 247794; CNA87744.1; -
 FTR: A31086; MUBPCP.
 DR InterPro: IPR002479; CW_binding.
 DR InterPro: IPR002053; Glycosyl_hydrolase_25.
 DR Pfam: PF01473; CW_binding.1; 4.
 DR Pfam: PF01183; Glyco_hydro.25; 1.
 DR ProDom: PD004620; Glycosyl_hydrolase_25; 1.
 DR PROSITE: PS00953; GLYCOSYL_HYDROL_F25; 1.
 DR KW Hydrolase; Glycosidase; Bacteriolytic enzyme; Repeat.
 FT ACT_SITE 10 10 BY SIMILARITY.
 FT ACT_SITE 37 37 BY SIMILARITY.

FT DOMAIN 201 324 6 X 20 AA TANDEM REPEATS.
 FT REPEAT 201 221 1.
 FT REPEAT 222 242 2.
 FT REPEAT 243 262 3.
 FT REPEAT 263 282 4.
 FT REPEAT 283 304 5.
 FT REPEAT 305 324 6.
 FT CONFLICT 44 44 P -> R (IN REF. 2).
 SQ SEQUENCE 339 AA; 39189 MW; E2947566AC676B05 CRC64;

Query Match 7.7%; Score 110.5; DB 1; Length 339;
 Best Local Similarity 22.4%; Pred. No. 0.14;
 Matches 51; Conservative 43; Mismatches 87; Indels 47; Gaps 11;

QY 58 YPEIFTKQVSTGVAIDPSQKITE-----WSAKYFNGTNILHRSGYVLKADGGY--YN 110
 Db 120 YKPIIYKYKPFTHNDVQOILAQFPNSIWIAGY--GLNDGTANFEYFPMGIRHWQYS 177
 QY 111 LGTTTKQFLPI-----QLTGEKKQGNF-FVKGNQGN-----YFYDLA 149
 Db 178 SNPFDRNVLDDDEDEKPKTAGTWKQSKGWHFRNNGSFYKWKKEIGGVWYFDSKG 237
 QY 150 NMVKNFTFIEDSVGNWYFFQDQDKMVENKHFVDVDSYGERGTFFFLKNGVSFRGLVQTDN 209
 Db 238 YCITSEWLKDN-EKWTYLDKNGAMVTG--WVLVGS-----EWYMDSDGAMVTGWVKYKN 289
 QY 210 GTTYFND-YGKVRNQTINAGMIYTLDECKLIKASYNDAEYPTST 256
 Db 290 NWYMTNERNGNMVSNEFKSGKGYFMTNGEL-----ADNPSFT 329

RESULT 15

HAPT_VIBCH STANDARD; PRT: 609 AA.
 ID HAPT_VIBCH
 AC P24153; Q5JX23;
 DT 01-MAR-1992 (Rel. 21, Created)
 DT 01-MAR-1992 (Rel. 21, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hemagglutinin/proteinase precursor (EC 3.4.24.-) (HA/protease) (Vibriolysin).
 GN HAP OR VCA0865.
 OS Vibrio cholerae.
 OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
 OX NCBI_TaxID=666;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=3083;
 RX MEDLINE=91258311; PubMed=20453361;
 RA Haese C.C., Finkelstein R.A.;
 RT "Cloning and nucleotide sequence of the Vibrio cholerae hemagglutinin/protease (HA/protease) gene and construction of an HA/protease-negative strain";
 RL J. Bacteriol. 173:3311-3317(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=EL TOR N16961 / SEROTYPE O1;
 RX MEDLINE=20406833; PubMed=10952301;
 RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L., Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A., Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D., Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragol I., Sellers P., McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O., Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C., Fraser C.M.;
 RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae";
 RL Nature 406:477-483(2000).
 RN [3]
 RP SEQUENCE OF 197-217.
 RC STRAIN=CLASSICAL CA401;
 RX MEDLINE=91071884; PubMed=2123831;
 RA Haese C.C., Finkelstein R.A.;

"Comparison of the *Vibrio cholerae* hemagglutinin/protease and the *Pseudomonas aeruginosa* elastase.";
RL Infect. Immun. 58:4011-4015(1990).
CC -!- FUNCTION: MAY PLAY A ROLE IN THE PATHOGENESIS OF CHOLERA. HAP
CC NICKS AND ACTIVATES THE A SUBUNIT OF CHOLERA ENTEROTOXIN AND
CC RELATED ENTEROTOXINS.
CC -!- COFACTOR: BINDS 1 ZINC ION.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M4 (ZINC METALLOPROTEASE);
CC ALSO KNOWN AS THE THERMOLYSIN SUBFAMILY.
CC -!- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-3 IS THE INITIATOR.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M59466; AAA27579.1; -.
CC EMBL: AE004414; AAF96763.1; -.
CC PIR: A42358; A42358.
CC HSP: P14756; 1E2M.
CC MEROPS: M04.003; -.
CC TIGR: VCA0865; -.
CC InterPro: IPR001570; Peptidase_M4.
CC InterPro: IPR000130; Zn_Metpeptidase.
CC Pfam: PF01447; Peptidase_M4; 1.
CC Pfam: PF02868; Peptidase_M4_C; 1.
CC PRINTS: PR00730; THERMOLYSIN.
CC PROSITE: PS00142; ZINC_PROTEASE; 1.
CC Hydrolase: Metalloprotease; Calcium, zinc; zymogen; signal;
KW Complete proteome.
FT SIGNAL 1 24
FT PROPEP 25 196
FT CHAIN 197 609
FT METAL 343 343
FT ACT_SITE 344 344
FT METAL 347 347
FT METAL 367 367
FT ACT_SITE 426 426
FT SEQUENCE 609 AA; 65891 MW; 6D003A589B956FD0 CRC64;

Query Match 7.4%; Score 106.5; DB 1; Length 609;
Best Local Similarity 23.1%; Pred. No. 0.59; 94; Indels 55; Gaps 12;
Matches 56; Conservative 37; Mismatches 94; Indels 55; Gaps 12;
QY 71 VAIDPSKITEWSAKYFNGSTNIIHRGSGYVLKADGGQYNNLCTTTKQLP-IQLTCEKQ 129
DB 285 VVFDMYQWLNTSPLTFLQTLTRVHVGNNYENAFWDMGAMTEGDTYRFPYLDVNSAHE 344
QY 130 GNEGFKVGNHGNY-----FYDLAG-----NWKNFTIEDSVGNWYFED 168
DB 345 VSHGFEQNSGLYRDMDSGGINAEAFSDIAGEAAEYPMRGNVDWIVGADIFKSSGGLRYFD 404
QY 169 Q---DGKWEV---NKEFVDVDSYGEKTY---PFL---KNGVSFREG-----LVQTD 208
DB 405 QPSRDRGSRDHSQSYSGIDVHSSGSFVRARYLLANKSGWNRKGFVFVAVANQLYWP 464
QY 209 NGTYFDNYG-----KMRNQTINAGAMIYTLDENGKLKASYNDAEYPTSDVGMQLDQ 264
DB 465 NST--FDQGGCGVYKAAQDLNFTADVAAFTVG--VNASGTT-----TPPVGKVLK 515
QY 265 NK 266
DB 516 GK 517

RESULT 16
SLAP_ACEKI STANDARD; PRT; 762 AA.
ID SLAP_ACEKI

P22258;
01-AUG-1991 (Rel. 19, Created)
01-AUG-1991 (Rel. 19, Last sequence update)
16-OCT-2001 (Rel. 40, Last annotation update)
DE Cell surface protein precursor (S-layer protein).
OS Acetogenium kivui
CC Bacteria; Firmicutes; Bacillus/Clostridium group;
CC Thermoanaerobacter group; Thermoanaerobacter.
OX NCBI_TaxID=2325;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=DSM 2030;
RX MEDLINE=90036724; PubMed=2681162;
RA Peters J., Peters M., Lottspeich F., Baumeister W.;
RT "S-layer protein gene of *Acetogenium kivui*: cloning and expression in
RT *Escherichia coli* and determination of the nucleotide sequence.";
RL J. Bacteriol. 171:6307-6315(1989).
RN [2]
RP PARTIAL SEQUENCE, AND CARBOHYDRATE-LINKAGE SITES.
RC STRAIN=DSM 2030;
RX MEDLINE=92281680; PubMed=1596358;
RA Peters J., Rudolf S., Oschkinat H., Mengele R., Sumper M.,
RA Kellermann J., Lottspeich F., Baumeister W.;
RT "Evidence for tyrosine-linked glycosaminoglycan in a bacterial
RT surface protein.";
RL Biol. Chem. Hoppe-Seyler 373:171-176(1992).
RN [3]
RP DOMAINS.
RX MEDLINE=94156823; PubMed=8113161;
RA Lupas A., Engelhardt H., Peters J., Santarius U., Volker S.,
RA Baumeister W.;
RT "Domain structure of the *Acetogenium kivui* surface layer revealed by
RT electron crystallography and sequence analysis.";
RL J. Bacteriol. 176:1224-1233(1994).
CC -!- FUNCTION: THE S-LAYER IS A PARACRYSTALLINE MONO-LAYERED ASSEMBLY
CC OF PROTEINS WHICH COAT THE SURFACE OF BACTERIA.
CC -!- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A
CC S-LAYER WITH HEXAGONAL SYMMETRY.
CC -!- PTM: THE CARBOHYDRATE CONTENT OF THIS PROTEIN IS ABOUT 8% WHICH
CC CORRESPOND TO ABOUT 40 TO 50 SUGAR MOLECULES PER MONOMER. O-LINKED
CC GLYCANS CONSIST OF GLC, GALNAc AND GLCNAC.
CC -!- SIMILARITY: CONTAINS 3 S-LAYER HOMOLGY (SLH) DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M31069; AAA21930.1; -.
CC PIR: A34355; A34355.
CC InterPro: IPR001119; SLH.
CC Pfam: PF00395; SLH; 2.
CC PROSITE: PS01072; SLH_DOMAIN; 2.
KW Signal; Glycoprotein; Repeat; Cell wall; S-layer.
FT SIGNAL 1 26
FT CHAIN 27 762
FT DOMAIN 30 94
FT SLH 1.
FT DOMAIN 95 145
FT SLH 2.
FT DOMAIN 146 204
FT SLH 3.
FT DOMAIN 473 479
FT SER/THR-RICH.
FT DOMAIN 625 630
FT SER/THR-RICH.
FT CARBOHYD 297 297
FT O-LINKED (GLC. . .).
FT CARBOHYD 516 516
FT O-LINKED (GLC. . .).
FT CARBOHYD 520 520
FT O-LINKED (GLC. . .).
FT CARBOHYD 632 632
FT O-LINKED (GLC. . .).
SQ SEQUENCE 762 AA; 82785 MW; 34EC9C784DECA67E CRC64;

Query Match 7.4%; Score 106.5; DB 1; Length 762;
Best Local Similarity 24.1%; Pred. No. 0.77;


```
DR pfam: PF00041; fn3; 2.
DR SMART; SM00060; FN3; 2.
KW Hydrolase; Glycosidase; Carbohydrate metabolism; Signal; Repeat;
FT Multifunctional enzyme.
FT SIGNAL 1 31
FT CHAIN 32 1481
FT DOMAIN 926 1015
FT DOMAIN 1158 1250
FT ACT_SITE 628 628
FT ACT_SITE 657 657
FT ACT_SITE 734 734
SQ SEQUENCE 1481 AA; 166362 MW; CB2635960539CALD CRC64;

Query Match 7.0%; Score 100; DB 1; Length 1481;
Best Local Similarity 26.4%; Pred. No. 5.4;
Matches 75; Conservative 35; Mismatches 98; Indels 76; Gaps 19;

QY 7 KEVATVTRVDDRCNVKWDALINNNL--YVVNTIG--GGYQKKYGGAFDLKLOKLYPEIF 62
DB 990 KIVSYTAVDSGN---ESALSNEVAYPAFISGAGNNQVDTIVGVNNPVEVIAELW 1046
QY 63 TKKQVSTGVADIPSQKITEWSAKYFNGTILHRGSGYVLKADGGQYINLTGTTKQFLPTQ 122
DB 1047 AE-----GLTDKPGQ-----GENMIAQ-LGYRYIGSGQ----DATENKVEGVE 1085
QY 123 LTGEKKQNGEGFVKNGDGNVYFYDLAGNAYKVTFTEDSVGNW-----YFF 167
DB 1086 INKDTWVDARYV-GDSGNNDYI-----NAK--FVPMVGTWEYIMRFSNQGQDWTYTK 1137
QY 168 DDQDKAVENKHFVDVSYG-EGTYFFFLKN-GV-SFRGLG---VOTDNGTYFYDNKGMV 221
DB 1138 GPGKTDKAKQIVPVSNDVEPTALGLOQPGIESRVTLSNLSSTDVAYI-----1189
QY 222 RNOTINAGAMYI-TLDENKGLKASYNDAEYP-TSTDV--GKM 261
DB 1190 -----GYEIKSLSETGPFVKIATVADTVINYVDTDVNGKV 1226

RESULT 20
PIP_LACLC STANDARD; PRT; 1902 AA.
AC P16271;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE P1-type proteinase precursor (EC 3.4.21.-) (Wall-associated serine
proteinase).
GN P1TP.
OS Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).
OG Plasmid pW05.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Lactococcus.
OX NCBI_Taxid=1359;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-WG2;
RX MEDLINE=88149035; PubMed=3278687;
RA Kok J., Leenhouts K.J., Haandrikman A.J., Ledeboer A.M., Venema G.;
RT "Nucleotide sequence of the cell wall proteinase gene of
Streptococcus cremoris Wg2."
RL Appl. Environ. Microbiol. 54:231-238(1988).
CC -1- FUNCTION: PROTEASE WHICH BREAKS DOWN MILK PROTEINS DURING THE
GROWTH OF THE BACTERIA ON MILK.
CC -1- CATALYTIC ACTIVITY: ENDOPEPTIDASE ACTIVITY WITH VERY BROAD
SPECIFICITY, ALTHOUGH SOME SUBSITE PREFERENCE HAVE BEEN NOTED,
E.G. LARGE HYDROPHOBIC RESIDUES IN THE P1 AND P4 POSITIONS, AND
PRO IN THE P2 POSITION. BEST KNOWN FOR ITS ACTION ON CASEINS,
ALTHOUGH IT HAS BEEN SHOWN TO HYDROLYSE HEMOGLOBIN AND OXIDIZED
INSULIN B-CHAIN.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Cell wall.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
SUBTILASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M24767; AAA17677.1; -.
CC HSP; P00782; I501.
CC MEROPS; S08.019; -.
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR003137; PA.
DR InterPro; IPR000209; Peptidase_S8.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR Pfam; PF02225; PA; 1.
DR Pfam; PF00082; Peptidase_S8; 3.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00136; SUBTILASE_ASP; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; 1.
KW Hydrolase; Serine protease; Cell wall; zymogen; Signal; Plasmid;
KW Transmembrane.
FT SIGNAL 1 33
FT PROPEP 34 187
FT CHAIN 188 1902
FT DOMAIN 188 1876
FT TRANSMEM 1877 1895
FT DOMAIN 1896 1902
FT ACT_SITE 217 217
FT ACT_SITE 281 281
FT ACT_SITE 620 620
FT DOMAIN 1867 1872
SQ SEQUENCE 1902 AA; 199910 MW; 2901C7F19B2E5D0B CRC64;

Query Match 7.0%; Score 99.5; DB 1; Length 1902;
Best Local Similarity 23.8%; Pred. No. 8;
Matches 67; Conservative 44; Mismatches 82; Indels 89; Gaps 17;

QY 8 EVATVTRVDDRCNVKWDALINNNLVVN--TIGGGEYQKKYGGAFDLKLOK----- 57
DB 664 ENNTAQPIID-----INYNVIVSPRQAGLVDPK---AAIDALEKNPSTVVAEN 711
QY 58 -YPEIFTKKQVSTGVADIPSQKI-----TEWSAKYFNGTN-----ILH- 94
DB 712 GYPVELKDFSTST---DKTFKLTFTNSTHETLYQMSNTDTNAVYTSATDPNSGVLYD 767
QY 95 -RGSYVLKADGGQYINLTGTTK-QF-LPIQLTGKKKGNGEYV--KNDGNYFYFDLAG 149
DB 768 KKIDGAIAKAGSNITVPASAKTAQIEFTLSLPKSPQDQGFVGFGLNFKGSDGSRNLPMYG 827
QY 150 NMVKNFTFTEDSVGNVYFFDQDQKAVENKHFVDV--SYGKGTTFELKN---GVSPRGVLV 205
DB 828 -----PFGDW-----NDGKIYVDSLNGITYSPAGNFGTVPLLTNNKNTGTQYGGW 873
QY 206 QTDNGTYFYFDNYGKAVRQNTINAGAMYITLDENGKLIKASYN 247
DB 874 TDADG-----NOTVDDQAIASFSSDKN-----ALYN 898

RESULT 21
P2P_LACLC STANDARD; PRT; 1902 AA.
ID P2P_LACLC
AC P15293;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE P1I-type proteinase precursor (EC 3.4.21.96) (LactococpepIn) (Cell wall-
associated serine proteinase) (LP151).
```



```

Db 226 VPCQYXNFRVLDAAEYDLSGETYDFASIMH--YARNTFS-----RGIWLDITILPRKD 275
QY 55 --OKLYPEITFKKQVSTGWAID-----PSQKIT-----EWAKYFNGTNGILH 94
Db 276 PESGIRPEIQGRKHLSEGDIIQANLLYKPCSGCRTLLSTGNFSSPEWPGQY-----327
QY 95 RSGGYVLKADGGQ--YVNLGTTTKQFLPQLTGEKKQGNEG---FVKNGDNGYFYDILA 148
Db 328 -----DGDQTCVMRSVTPGETISLQFTGLVSGDGCWYNTLEVRDGHWRHSPLL 378
QY 149 GNMVKNFTFIDSYGNMYFFDQDQKMYENKHFVDVDSYGEKGTFFFLKNGVSFSGGLVQTD 208
Db 379 GRFCGASLPDPILSS-----DSRLWIELKSSAHYSRGRFAANVEAICGGHIERE 427
QY 209 NGTYFYDFNY 217
Db 428 SGTILQSPNY 436

RESULT 25
SFMD_ECOLI STANDARD; PRT; 867 AA.
AC P77468; P77133;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Outer membrane usher protein sfmd precursor.
GN SFMD OR B0532.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-X12 / MGL655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RA "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RA Chung E., Allen E., Araujo R., Aparicio A., Davis K., Duncan M.,
RA Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O., Lew H., Lin D.,
RA Namath A., Oefner F., Roberts D., Schramm S., Davis R.W.;
RT "Sequence of minutes 4-25 of Escherichia coli.";
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: INVOLVED IN THE EXPORT AND ASSEMBLY OF SPWA FIMBRIAL
CC SUBUNITS ACROSS THE OUTER MEMBRANE (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane
CC (by similarity).
CC -1- SIMILARITY: BELONGS TO THE FIMBRIAL EXPORT USHER FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE000159; AAC73634.1; -.
CC EMBL; U82598; AAB40730.1; -.
CC EMBL; U82664; AAB40285.1; -.
CC EcoGene; EG13883; sfmd.
CC InterPro; IPR000015; Fimb_usher.
CC Pfam; PF00577; Usher; 1.
CC PROSITE; PS01151; FIMBRIAL_USHER; 1.
KW Outer membrane; Transmembrane; Fimbria; Transport; Signal;
KW Complete proteome.
FT SIGNAL 1 35 POTENTIAL.

```

```

FT CHAIN 36 867 OUTER MEMBRANE USHER PROTEIN SFMD.
FT DISULFID 840 862 POTENTIAL..
SQ SEQUENCE 867 AA; 95677 MW; DF8591D0B6C4205A CRC64;

Query Match
Best Local Similarity 21.9%; Score 96; DB 1; Length 867;
Matches 68; Conservative 53; Mismatches 118; Indels 72; Gaps 17;

QY 6 GKEVATVTKVDRGN-----VWK-DAIINN--LYVNTTIGGEYKKG--- 48
Db 48 GTSVADLSRF-ERGNHQPAQVYRVDLWRNDEFISQDIVFESTTENTGD---KSGGLAPC 103
QY 49 ---AFIDKL---OKLYPEIF---TKQVSTGVADPSQKITSESAKYFNGT---NILHR 95
Db 104 FNOVLDERIGLNSAPPELAQQNNKCIINLLKAVPDATINFDFAAMRLNITIPQIALSS 163
QY 96 GSGYVLKADGGQYVNLGTTTKQFLPQLTGEKKQGNEGFVKNGDNGYFYDILAGNMVKNY 155
Db 164 AHGYIPPEE-----WDBGIPALLNLYNFTGNRG--NGND-SYFSELSGINIGPW 210
QY 156 FTDSVGNMYFFDQDQKMYENKHFVD-----VDSYGEKGTFFFLKNGVSFR 201
Db 211 RLKNN-GSNYFRGNGYHSEQNNNIGTWTQRAIIPLKSELVMDGNTGSDIF--DGVGFR 267
QY 202 GGIWQTDNCTYFYFDNGYKMYRNOTINAGAMIVTLDENGLIKASYNDAEY-----PTS 255
Db 268 GVKLYSSDNNYDPSQQGFAPTGVGTARTAAQTIRONGFIITQSYVSPGAFITDLHPTS 327
QY 256 T--DVGKMLDQ 264
Db 328 SNGDLVDVTIDE 338

```

Search completed: August 12, 2002, 08:08:48
Job time: 220 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 12, 2002, 08:04:48 ; Search time 26.51 Seconds

(without alignments)

1742.348 Million cell updates/sec

Title: US-09-995-749a-2_COPY_1515_1781

Perfect score: 1430

Sequence: 1 IYNLPKEVATRVDRGN.....SDAEYPTSDVGKMLDQNK 267

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

1: SP TREMBL_19:*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phase.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_virus.*
16: sp_bacteriap.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	582.5	40.7	2057	Q9RE05	Q9re05 leuconostoc
2	540.5	37.8	1477	Q9L466	Q9L466 leuconostoc
3	540.5	37.8	1508	O52224	O52224 leuconostoc
4	533.5	37.3	1508	Q9EZH5	Q9ezh5 leuconostoc
5	531	37.1	1338	Q9WXJ4	Q9wxj4 streptococ
6	500	35.0	1577	O54178	O54178 streptococ
7	497	34.8	1527	Q9ZAR4	Q9zar4 leuconostoc
8	465	32.5	1575	Q9LCH3	Q9lch3 streptococ
9	458.5	32.1	1512	Q9WXJ5	Q9wxj5 streptococ
10	448	31.3	1449	O55264	O55264 streptococ
11	446	31.2	1449	O68542	O68542 streptococ
12	445.5	31.2	1390	O69385	O69385 streptococ
13	445.5	31.2	1455	O69382	O69382 streptococ
14	445.5	31.2	1455	O69391	O69391 streptococ
15	445.5	31.2	1455	O69397	O69397 streptococ
16	440.5	30.8	1590	Q55263	Q55263 streptococ

17	439.5	30.7	1590	2	Q59983	Q59983 streptococ
18	439	30.7	1290	2	Q48756	Q48756 leuconostoc
19	438.5	30.7	1455	2	O69388	O69388 streptococ
20	437	30.6	1599	2	Q00599	Q00599 streptococ
21	428.5	30.0	1518	2	Q00600	Q00600 streptococ
22	418	29.2	1577	2	O55265	O55265 streptococ
23	169	11.8	2817	16	Q97K42	Q97k42 clostridium
24	156	10.9	316	9	Q38658	Q38658 bacterioph
25	134.5	10.8	563	2	Q54447	Q54447 streptococ
26	147.5	10.3	332	16	Q97RW9	Q97rw9 streptococ
27	145.5	10.2	482	16	Q9CDJ7	Q9cdj7 lactococcus
28	143	10.0	302	2	Q9S112	Q9s112 streptococ
29	142.5	10.0	330	2	O55228	O55228 streptococ
30	142	9.9	318	9	Q94M29	Q94m29 streptococ
31	142	9.9	483	16	Q9CDJ6	Q9cdj6 lactococcus
32	138	9.7	300	2	Q9L412	Q9l412 streptococ
33	137	9.6	393	2	Q9ZFN1	Q9zfn1 peptostrept
34	136.5	9.5	601	2	Q45821	Q45821 clostridium
35	134.5	9.4	614	2	Q9AH78	Q9ah78 streptococ
36	133	9.3	300	2	Q9L410	Q9l410 streptococ
37	133	9.3	302	2	Q9R315	Q9r315 streptococ
38	133	9.3	313	2	Q9AG23	Q9ag23 streptococ
39	132	9.2	302	2	Q9R437	Q9r437 streptococ
40	132	9.2	302	2	Q9R3M2	Q9r3m2 streptococ
41	132	9.2	302	2	Q9R3M1	Q9r3m1 streptococ
42	132	9.2	312	2	Q9AG25	Q9ag25 streptococ
43	132	9.2	313	2	Q9AG24	Q9ag24 streptococ
44	131	9.2	300	2	Q9L414	Q9l414 streptococ
45	131	9.2	300	2	Q9L411	Q9l411 streptococ
46	131	9.2	300	2	Q9L409	Q9l409 streptococ
47	131	9.2	302	2	Q9S113	Q9s113 streptococ
48	131	9.2	316	2	O53275	O53275 streptococ
49	130	9.1	300	2	Q9L408	Q9l408 streptococ
50	129.5	9.1	302	2	Q9S111	Q9s111 streptococ
51	129	9.0	300	2	Q9L413	Q9l413 streptococ
52	128.5	9.0	340	16	Q97S14	Q97s14 streptococ
53	128	9.0	583	2	Q45822	Q45822 clostridium
54	125.5	8.8	752	2	Q9KK30	Q9kk30 streptococ
55	125	8.7	627	16	Q9KGZ1	Q9kgz1 streptococ
56	124.5	8.7	649	2	Q03706	Q03706 clostridium
57	123.5	8.6	621	16	Q97NB5	Q97nb5 streptococ
58	123	8.6	627	2	Q9AJT3	Q9ajt3 streptococ
59	121	8.5	283	9	Q9AF60	Q9af60 streptococ
60	120.5	8.4	606	2	O66380	O66380 erysipeloth
61	120.5	8.4	2364	2	Q46342	Q46342 clostridium
62	118.5	8.3	684	2	Q9KK17	Q9kk17 streptococ
63	118	8.3	627	2	Q9E0R5	Q9eur5 streptococ
64	117	8.2	730	2	Q9KK47	Q9kk47 streptococ
65	116.5	8.1	296	9	O03979	O03979 bacterioph
66	116.5	8.1	590	2	Q45819	Q45819 clostridium
67	116.5	8.1	667	2	Q9KK28	Q9kk28 streptococ
68	116.5	8.1	708	2	O05166	O05166 streptococ
69	116.5	8.1	2367	2	Q9F931	Q9f931 clostridium
70	116.5	8.1	2367	2	Q46034	Q46034 clostridium
71	116	8.1	2367	2	Q9EXR0	Q9exr0 clostridium
72	115.5	8.1	635	2	Q45820	Q45820 clostridium
73	115.5	8.1	752	2	Q9KK33	Q9kk33 streptococ
74	115	8.0	680	2	Q9KK11	Q9kk11 streptococ
75	115	8.0	681	2	Q9KK13	Q9kk13 streptococ
76	115	8.0	2399	16	Q9ZKS9	Q9zks9 helicobacte
77	114.5	8.0	539	2	O33741	O33741 streptococ
78	114.5	8.0	655	2	Q9KK50	Q9kk50 streptococ
79	114.5	8.0	663	2	O30874	O30874 streptococ
80	114.5	8.0	701	2	Q9RQF5	Q9rqt5 streptococ
81	114.5	8.0	701	2	Q9RQF5	Q9rqt5 streptococ
82	114	8.0	696	2	Q9KK48	Q9kk48 streptococ
83	113.5	7.9	832	3	Q9KK14	Q9kk14 streptococ
84	113.5	7.9	929	2	Q9KK19	Q9kk19 streptococ
85	113.5	7.9	929	2	Q9KAY5	Q9kay5 streptococ
86	113.5	7.9	2178	2	Q46149	Q46149 clostridium
87	113	7.9	261	17	Q9V145	Q9v145 pyrococcus
88	113	7.9	695	2	Q9KK51	Q9kk51 streptococ
89	113	7.9	696	2	Q9KK32	Q9kk32 streptococ

RT "Cloning and sequencing of a gene coding for an extracellular
 RT dextranase (DSRB) from *Leuconostoc mesenteroides* NRRL B-1299
 RT synthesizing only a (1-6) glucan."
 RL FEMS Microbiol. Lett. 0:0-0(1998).
 DR EMBL; AF030129; AAB95453.1; .
 DR InterPro; IPR002479; CW_binding.
 DR InterPro; IPR003318; Glyco_hydro_70.
 DR Pfam; PF01473; CW_binding_1; 14.
 DR Pfam; PF02324; Glyco_hydro_70; 1.
 DR Transferrase; Glycosyltransferase.
 KW
 SQ SEQUENCE 1508 AA; 168511 MW; E70CECB57A70D1F0 CRC64;

Query Match 37.8%; Score 540.5; DB 2; Length 1508;
 Best Local Similarity 43.0%; Pred. No. 7e-28; Indels 35; Gaps 10;
 Matches 119; Conservative 47; Mismatches 76;

QY 1 IYNLPGEVATVTRVDDRGNYWKDAIINNLYVNTIGGQYKKGAGFLDKLQKLYPE 60
 DB 1011 IYNLPGEVATVTRVDDRGNYWKDAIINNLYVNTIGGQYKKGAGFLDKLQKLYPE 1070
 QY 61 IFTKQVSTGVAIDPSOKITEWSAKYFNGTILHRGSGYVLKADG-GQYVNL--GTTTKQ 117
 DB 1071 LFEKQISTGLPDPSPKITEWSKIFNGSIOGKAGYVLKDSGTQYKYKVTNNNRD 1130
 QY 118 FLPIQLTGEKKQNEGVKNGDNGYFYDLAGNNVKNFTIEDSVGNWYFFDQDGKVENK 177
 DB 1131 FLPKQLTDLSE--TGFVRDNGVY-YTLGSLARNTFIQDDNGNYYFDSGHLVWG- 1186
 QY 178 HFVDVDSYGEKGYFFLKNKNGSVFRGLVQ-TDNGTYFYDNGYKGMVRNQT- 226
 DB 1187 -FQINNH---HYFFLPNGIELVQSLQADGSTIYFDQKGRQVFNQYITDQGTAYVF 1241
 QY 227 -NAGAMIVT---LD-----ENGLIKASYNDAE 251
 DB 1242 QNDGTMTVSGTEIDGHKQYFKNGTQVKGFVSQTD 1278

RESULT 4
 Q9EZH5 PRELIMINARY; PRT; 1508 AA.
 AC Q9EZH5;
 DT 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE DEXTRANSUCRASE DSRB742.
 GN DSRB742.
 OS *Leuconostoc mesenteroides*.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
 OC *Leuconostoc*.
 OX NCBI_TaxID=1245;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-B-742CB;
 RA Kim H.-S., Kim D., Ryu H.-J., Robyt J.F.;
 RL "Leuconostoc mesenteroides B-742CB, a dextranucrase gene."
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF294469; AAG38021.1; .
 DR InterPro; IPR002479; CW_binding.
 DR InterPro; IPR003318; Glyco_hydro_70.
 DR Pfam; PF01473; CW_binding_1; 14.
 DR Pfam; PF02324; Glyco_hydro_70; 1.
 SQ SEQUENCE 1508 AA; 168542 MW; E2FCFA0F87AE4F3A CRC64;

Query Match 37.3%; Score 533.5; DB 2; Length 1508;
 Best Local Similarity 42.6%; Pred. No. 2e-27;
 Matches 118; Conservative 47; Mismatches 77; Indels 35; Gaps 10;

QY 1 IYNLPGEVATVTRVDDRGNYWKDAIINNLYVNTIGGQYKKGAGFLDKLQKLYPE 60
 DB 1011 IYNLPGEVATVTRVDDRGNYWKDAIINNLYVNTIGGQYKKGAGFLDKLQKLYPE 1070

QY 61 IFTKQVSTGVAIDPSOKITEWSAKYFNGTILHRGSGYVLKADG-GQYVNL--GTTTKQ 117
 DB 1071 LFEKQISTGLPDPSPKITEWSKIFNGSIOGKAGYVLKDSGTQYKYKVTNNNRD 1130
 QY 118 FLPIQLTGEKKQNEGVKNGDNGYFYDLAGNNVKNFTIEDSVGNWYFFDQDGKVENK 177
 DB 1131 FLPKQLTDLSE--TGFVRDNGVY-YTLGSLARNTFIQDDNGNYYFDSGHLVWG- 1186
 QY 178 HFVDVDSYGEKGYFFLKNKNGSVFRGLVQ-TDNGTYFYDNGYKGMVRNQT- 226
 DB 1187 -FQINNH---HYFFLPNGIELVQSLQADGSTIYFDQKGRQVFNQYITDQGTAYVF 1241
 QY 227 -NAGAMIVT---LD-----ENGLIKASYNDAE 251
 DB 1242 QNDGTMTVSGTEIDGHKQYFKNGTQVKGFVSQTD 1278

RESULT 5
 Q9MAJ4 PRELIMINARY; PRT; 1338 AA.
 AC Q9MAJ4;
 DT 01-NOV-1999 (Tremblrel. 12, Created)
 DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE GTF-S.
 GN *Streptococcus criceti*.
 OS *Streptococcus criceti*.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 OC *Streptococcus*.
 OX NCBI_TaxID=1333;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-HS-6;
 RA Inoue M., Fukui K., Miyagi A.;
 RL "S. cricetus glucosyltransferase (gifs and gtf) genes."
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB026123; BAA77336.1; .
 DR InterPro; IPR002479; CW_binding.
 DR InterPro; IPR003318; Glyco_hydro_70.
 DR Pfam; PF01473; CW_binding_1; 10.
 DR Pfam; PF02324; Glyco_hydro_70; 1.
 KW Plasmid.
 SQ SEQUENCE 1338 AA; 148558 MW; 0A90C8E10F15D99B CRC64;

Query Match 37.1%; Score 531; DB 2; Length 1338;
 Best Local Similarity 41.2%; Pred. No. 2.6e-27;
 Matches 128; Conservative 33; Mismatches 98; Indels 52; Gaps 9;

QY 1 IYNLPGEVATVTRVDDRGNYWKDAIINNLYVNTIGGQYKKGAGFLDKLQKLYPE 60
 DB 893 IYNLPGEVATVTRVDDRGNYWKDAIINNLYVNTIGGQYKKGAGFLDKLQKLYPS 952
 QY 61 IFTKQVSTGVAIDPSOKITEWSAKYFNGTILHRGSGYVLKADGQYVNLGTTTKQFLP 120
 DB 953 LFTTKQISTGKPIDSPKITEWSAKYFNGSILGRGAKYVL-SDNNKYLMG-AGQFFLP 1010
 QY 121 IQLT---GEKKQNEGVKNGDNGYFYDLAGNNVKNFTIEDSVGNWYFFDQDGKMY--- 174
 DB 1011 TNLNNTYGPQAPANGPANGTSKNGGIIHYI-DNNGQEVKNQF-KEIAGSWYYPDANGKMATGQ 1068
 QY 175 -----ENKHFVDVDSYGEKGYFFLKN 196
 DB 1069 TKIGNTYFLFMPNGKQLKEGVWYDGGKRAYIYDDNGRWNTNKGFEVKVNGQDKWRYFNGD 1128
 QY 197 GVSFRGGLVQDNGTYFYDNGYKGMVRNQTINAGAMIVTLDEN-CKLIKASYNDAEYPTS 255
 DB 1129 G-SIAGVLSLDNRLTYFDAYGYQVKGQTLTNGKTYSFDAEAGDLITG--NTFSPFNN 1185
 QY 256 TDVSKMLDQNK 266
 DB 1186 QGAEALGDNQ 1196

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
OC Leuconostoc.

OX NCBI_TaxID=1245;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=NRRL B-512-F;

RA Bhattacharya R., Singh D.K.S.;

RT "Cloning and Molecular Characterization of Dextranucrase Gene from

Leuconostoc mesenteroides NRRL B-512F.";

RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.

DR EMBL: U81374; AAD10952.1; -

DR InterPro: IPR002479; CW_binding.

DR InterPro: IPR003318; Glyco_hydro_70.

DR Pfam: PF01473; CW_binding_1; 16.

DR Pfam: PF02324; Glyco_hydro_70; 1.

SQ SEQUENCE 1527 AA; 169709 MW; 1DFAFA237C743398 CRC64;

Query Match 34.8%; Score 497; DB 2; Length 1527;

Best Local Similarity 42.1%; Pred. No. 5.6e-25;

Matches 118; Conservative 44; Mismatches 78; Indels 40; Gaps 11;

QY 1 IYNLPKEVATVTRVDDRGVWKDAIINNLYVVTIGSGEYQKYGAGFLDKLQLYPE 60

DB 1030 IYNLPKEVATVTRINSEFGDDDDTSDIDNALYVQSRGGGQIQEMIGAGFLDLQALYPS 1089

QY 61 IFTKKQSTGVAIDPSOKITWESAKYFNGTILHRGSGYVLKADG-QQYVNLGTTTK--Q 117

DB 1090 LFKVNOISTGVPIIDGSKVITWAAKYFNGSNIQKGAGYVLKDNKSNKYFKVWSTEDGD 1149

QY 118 FLPTLIGCKEQGNEGFVKGNQNYFYFDLAGNWKNTFIEDSYGNWYFFDQDGKMYE-- 175

DB 1150 YLPKQLTNDLSE--TGFTHDDKGIY-YTLGSGYRAQNAFIQDDNNYFYFDKTHGLVTL 1206

QY 176 ---NKHFDVDSYGEKTYFFLKNKSVSPRGVLQVTDNGT-YFFDNYGKWRNQTI---NA 228

DB 1207 QKINNH-----TYFFLPNGELVSKFLONEDGTIVYFDKKGHQYFDQYITDQNG 1255

QY 229 GAMIYTLDENGKLK---ASYNDAEYPTSTDVGMKMLDQ 265.

DB 1256 NA--YFDDAGVMLKSLATIDGHQY-----FDQN 1284

RESULT 8

Q9LCH3

AC Q9LCH3 PRELIMINARY; PRT; 1575 AA.

DT 01-OCT-2000 (TrEMBLrel. 15, Created)

DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)

DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)

DE GLUCOSYLTRANSFERASE.

GN GTPR.

OS Streptococcus oralis.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;

OC Streptococcus.

OX NCBI_TaxID=1303;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC10557;

RX MEDLINE=20231779; PubMed=10768934;

RA Fujiwara T., Hoshino T., Sobue S., Hamada S.;

RT "Purification, characterization, and molecular analysis of the gene

encoding glucosyltransferase from Streptococcus oralis.";

RL Infect. Immun. 68:2475-2483(2000).

DR EMBL: AB025228; BAA95201.1; -

DR InterPro: IPR002479; CW_binding.

DR InterPro: IPR003318; Glyco_hydro_70.

DR Pfam: PF01473; CW_binding_1; 17.

DR Pfam: PF02324; Glyco_hydro_70; 1.

DR Transferase.

SQ SEQUENCE 1575 AA; 176792 MW; 772A26E4D7C2E543 CRC64;

RESULT 6

Q54178

ID Q54178 PRELIMINARY; PRT; 1577 AA.

AC Q54178; Q54247;

DT 01-NOV-1996 (TrEMBLrel. 01, Created)

DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)

DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)

DE GLUCOSYLTRANSFERASE.

GN GTFG.

OS Streptococcus gordonii challis.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;

OC Streptococcus.

OX NCBI_TaxID=28390;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CHALLIS;

RA MEDLINE=92276337; PubMed=1534326;

RA Vickerman M.M., Sulavik M.C., Clewell D.B.;

RT "Molecular analysis of Streptococcus gordonii glucosyltransferase

phase variants.";

RL Dev. Biol. Stand. 85:309-314(1995).

RN [2]

RP SEQUENCE OF 1-96 FROM N.A.

RC STRAIN=CHALLIS;

RX MEDLINE=92276337; PubMed=1534326;

RA Sulavik M.C., Tardif G., Clewell D.B.;

RT "Identification of a gene, igr, which regulates expression of

glucosyltransferase and influences the Spp phenotype of Streptococcus

gordonii Challis.";

RL J. Bacteriol. 174:3577-3586(1992).

DR EMBL: U12843; AAC43483.1; -

DR EMBL: M89776; AAA26969.1; -

DR InterPro: IPR002479; CW_binding.

DR InterPro: IPR003318; Glyco_hydro_70.

DR Pfam: PF01473; CW_binding_1; 18.

DR Pfam: PF02324; Glyco_hydro_70; 1.

DR Transferase.

SQ SEQUENCE 1577 AA; 177805 MW; 5AE0328DC5E08D18 CRC64;

Query Match 35.0%; Score 500; DB 2; Length 1577;

Best Local Similarity 46.2%; Pred. No. 3.7e-25;

Matches 110; Conservative 33; Mismatches 77; Indels 18; Gaps 7;

QY 1 IYNLPKEVATVTRVDDRGVWKDAIINNLYVVTIGSGEYQKYGAGFLDKLQLYP 59

DB 1012 IYNLPKEVATVTRVNNYTYREGAEIKELVANSKNTGTDYQKYGAGFLDELKAKYP 1071

QY 60 EIFTKKQSTGVAIDPSOKITWESAKYFNGTILHRGSGYVLKADGQYVNLGTTTKQFL 119

DB 1072 EIFERVQISNGOKMTTDEKITWESAKYFNGTILGKGYVVLKDWGSKYLSNKNGETAL 1131

QY 120 PIQLTGEKQKQNEGFVKGNQNYFYFDLAGNWKNTFIEDSYGNWYFFDQDGKMYENKHF 179

DB 1132 PKQLV--NKEASTGFVKDQNG-FKYSISGNQAKDTFIDENGNNYFYFDQGYLYTGARE 1188

QY 180 VVDVSGEKGTYFFLKNKSVSPRGVLQVTDNGT-YFPDNYGKWRNQTINAGAMIYTL 236

DB 1189 ID----GKQ--LYFMKNGQLRDALQEDENGQYDYDKTGAKVLNR-----YTTSD 1233

RESULT 7

Q9ZAR4

ID Q9ZAR4 PRELIMINARY; PRT; 1527 AA.

AC Q9ZAR4;

DT 01-MAY-1999 (TrEMBLrel. 10, Created)

DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)

DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)

DE DEXTRANSUCRASE.

GN DEX.

OS Leuconostoc mesenteroides.

Query Match 32.5%; Score 465; DB 2; Length 1575;
Best Local Similarity 41.2%; Pred. No. 7.9e-23;
Matches 110; Conservative 32; Mismatches 91; Indels 34; Gaps 8;

QY 1 IYNLPKEVATVTRVDDRGVWKDAIINNLLYVYVNT-IGGGEYQKYGGAFLDKLQKLYP 59
DB 1010 IYNLPKEVATVTRVYNNYTYREGAEIKEKLVANSKNTDTPQKYGGAFDELKAKYP 1069
QY 60 EIFTKKQVSTGVAIDPSOKITWMSAKYFNGNHLRSGYVLKADGGQYVNLGTTTKQFL 119
DB 1070 EITERVOISNGOKMTTDEKITKWSAKYFNGNHLRSGYVVLKADWASNDYLTNRNGEIVL 1129
QY 120 PQLQGEKKQKQNEGVKNDGNYFYDLAGNWKNTFLEDSVGNWYFFDQDKWVENKHF 179
DB 1130 PKQLV--NKNSTYGFVSDANGT-KEYTSYGVQAKNSFIQDENGWNYFFDKRGYLVTAHE 1186
QY 180 VVDVSYGKGTG-YFLKNGVSPRGGLVQTDNGT-YFYDNYGKWNQVNTINA----- 228
DB 1187 ID-----GRHVFELKNGIQLRDSIREDENGNQYYDQGAQVNLNRYTTDQGNWRYFD 1239
QY 229 --GAMYTYL-----DENGKLIK 243
DB 1240 AKGVNARGLVKIGDQQQFFDENGQYGVK 1266

RESULT 9
Q9WXJ5 PRELIMINARY; PRT; 1512 AA.
AC Q9WXJ5;
DT 01-NOV-1999 (TRENBLrel. 12, Created)
DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)
DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
DE GTF-S.
GN GTF.
OS Streptococcus criceti.
OC Plasmid pAM1.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1333;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-HS-6;
RA Inoue M., Fukui K., Miyagi A.;
RT "S.cricetus glucosyltransferase(gtfs and gtfS) genes.";
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB026123; BAN7237.1; .
DR InterPro; IPR002479; CW_binding.
DR DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 14.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Plasmid.
SQ SEQUENCE 1512 AA; 167145 MW; 4C03D9CBC601FC14 CRC64;

Query Match 32.1%; Score 458.5; DB 2; Length 1512;
Best Local Similarity 38.5%; Pred. No. 2e-22;
Matches 115; Conservative 38; Mismatches 95; Indels 51; Gaps 12;

QY 1 IYNLPKEVATVTRVDDRGVWKDAIINNLLYVYVNTIGG-EGYQKYGGAFLDKLQKLYP 59
DB 960 IYQLPGKEVYVTSRDVNGYRKVDQPMWKLANTKSSGKDFQAKYGGEFIAELQKQYP 1019
QY 60 EIFTKKQVSTGVAIDPSOKITWMSAKYFNGNHLRSGYVLKADG-QOYVNLGTTTKQF 118
DB 1020 EMTAKMISTGRPIDSSVLKENSAGYFNGTVLGRGTDFVLSDEGTGYFYV-NEKGF 1078
QY 119 LPQLQGEKKQKQNEGVKNDGNYFYDLAGNWKNTFLEDSVGNWYFFDQDKWVENKHF 178
DB 1079 LPAVLGSD-KEAKTGFYNDGKGMTYE-TTAGSQAKSDFV-TVAGNTYYFDYTGHWYVGN 1135
QY 179 FVDVDSYGEKGYFFELKNGVSPRGGLVQTDNG--YY----- 213
DB 1136 GINTK-----FYFLPNGVMUKDAYMEDDRGRSVYGTGVMYKGSRNNEWAMTDSKG 1189

QY 214 -----FDNYGKM-VRNQFTINAGAMITLDENGLIKASYNDAEYPTSTDVGKMLDQN 265
DB 1190 QLRFPHEDNYGFMVGLVTHGNVQY--DEESQGVKGFVTRKAGQT-----RYFDKN 1241

RESULT 10
Q55264 PRELIMINARY; PRT; 1449 AA.
AC Q55264;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
DE GLUCOSYLTRANSFERASE PRECURSOR.
GN GTF.
OS Streptococcus salivarius.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1304;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-95122197; Pubmed-7822030;
RA Simpson C.L., Giffard P.M., Jacques N.A.;
RT "Streptococcus salivarius ATCC 25975 possesses at least two genes
RT coding for primer-independent glucosyltransferases.";
RL Infect. Immun. 63:609-621(1995).
DR EMBL; L35495; AAC41412.1; .
DR InterPro; IPR002479; CW_binding.
DR DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 8.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Signal; Transferase.
FT SIGNAL 1 35 POTENTIAL.
FT CHAIN 36 1449 GLUCOSYLTRANSFERASE.
SQ SEQUENCE 1449 AA; 159984 MW; DD62F07306E86A46 CRC64;

Query Match 31.3%; Score 448; DB 2; Length 1449;
Best Local Similarity 40.1%; Pred. No. 9.7e-22;
Matches 113; Conservative 37; Mismatches 90; Indels 42; Gaps 12;

QY 1 IYNLPKEVATVTRVDDRGVWKDAIINNLLYVYVNT-IGGGEYQKYGGAFLDKLQKLYP 59
DB 996 IYNLPGEVATVTRVYNNYTYREGAEIKEKLVANSKNTDTPQKYGGAFDELKRLYP 1055
QY 60 EIFTKKQVSTGVAIDPSOKITWMSAKYFNGNHLRSGYVLKADGGQYVNLGTT-THQF 118
DB 1056 QIFDRVQISTGKRTTDEKITQWSAKYFNGNHLRSGYVVLKNGLNGY--GTNGSKVS 1113
QY 119 LPQLQGEKKQKQNEGVKNDGNT-----YFYDLAGNWKNTFIEDSVGNWYFFD 168
DB 1114 LP-KVVG-SNSTNGDNGQSGCKEFLFSVRYN-NGQYAKNAFIKNDGNVYFD 1170
QY 169 QGKVKVENKHFVDVDSYGEKGYFFELKNGVSPRGGLVQTDNG-TYFYDNYGKN----- 220
DB 1171 NSGRMAVGEKTDID---GKQ--YFFLANGVQLDGRYQNRGRQVFEYDQNGVLANQKQD 1224
QY 221 -----VRNQFTINAGAMIT-LDENGLIKASYNDAE 251
DB 1225 PKPDNNNASGRNQFVQIGNVWYDNGKRVTHGNQNG 1266

RESULT 11
O68542 PRELIMINARY; PRT; 1449 AA.
AC O68542;
DT 01-AUG-1998 (TRENBLrel. 07, Created)
DT 01-AUG-1998 (TRENBLrel. 07, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE GLUCOSYLTRANSFERASE N (FRAGMENT).
GN GTF.
OS Streptococcus salivarius.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;

```
OC Streptococcus.
OX NCBI_TaxID=1304;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-V1477;
RA Jaffe R.I.;
RL "Streptococcus salivarius V1477 gtfN.";
RE Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF049609; AAC05156.1; -.
DR InterPro; IPR002479; CW_binding_70.
DR Pfam; PF01473; Glyco_hydro_70.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Transferase.
FT NON_TER
FT 1449
FT SEQUENCE 1449 AA; 159895 MW; 0700F6D7484718FB CRC64;

Query Match      31.2%; Score 446; DB 2; Length 1449;
Best Local Similarity 40.1%; Pred. No. 1.3e-21;
Matches 113; Conservative 37; Mismatches 90; Indels 42; Gaps 12;

QY 1 IYNLPKEVATVTRVDDRGVNVKDAIINNLLYVNT-IGGGEYQKKYGGAFDLKQLKLYP 59
DB 996 IYNLPGEVVTATRVNNGYGETKGAIIHSLXAATRTFGNDYQKGIGGAFDLKRLYP 1055
QY 60 EIFTKKQVSTGVAIDPSOKITEWSAKYFNGTNIHRGSGYVLKADGGQYVNLGTTT-KQF 118
DB 1056 QIFDRVQISTGRMTTDEKITKWSAKYMGNTNLDRGSEYVLKNGLYY--GTNGGKVS 1113
QY 119 LPQLTGEKKQKNEGFEVKGNDG-----YFYDLAGNMVKNFTIEDSVGNWYFFD 168
DB 1114 LP-KVVG-SNQSTNGNDQNGDSGKFEKRLFSRYRYN-NGQYAKNAFTKDNQDNVYFFD 1170
QY 169 QDGKMYENKHEFVDVDSYGEKGYFELKNGVSRGGLVQTDNG-FYFVDNYGKMV----- 221
DB 1171 NSGRMVGKTDI----GKQ--YFFANGVQLKDGVTQRNGQVFIYDONGVLSANGKOD 1224
QY 222 -----RNQTNAGAMIYV-LDENGKLIKASYNSDAE 251
DB 1225 PKDNNNTSGRNQFVQIGNVWAYVDGNGKRVIGHQNINGO 1266

RESULT 12
ID O69385 PRELIMINARY; PRT; 1390 AA.
AC O69385;
DT 01-AUG-1998 (T-EMBLrel. 07, Created)
DT 01-AUG-1998 (T-EMBLrel. 07, Last sequence update)
DE 01-JUN-2001 (T-EMBLrel. 17, Last annotation update)
DE GLUCOSYLTRANSFERASE-SI.
GN GTPC.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-WT8148;
RX MEDLINE-98231643; PubMed-9570124;
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,
RA Kimura S., Hamada S.;
RT "Molecular analyses of glucosyltransferase genes among strains of
RT Streptococcus mutans.";
RL FEMS Microbiol. Lett. 161:331-336(1998).
DR EMBL; D88652; BAA26106.1; -.
DR InterPro; IPR002479; CW_binding_70.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 7.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Transferase.
KW SEQUENCE 1390 AA; 155375 MW; 8847E4956EF05E9F CRC64;

Query Match      31.2%; Score 445.5; DB 2; Length 1455;
Best Local Similarity 39.9%; Pred. No. 1.4e-21;
Matches 115; Conservative 32; Mismatches 96; Indels 45; Gaps 13;

QY 1 IYNLPKEVATVTRVDDRGVNVKDAIINNLLYVNV-TIGGGEYQKKYGGAFDLKQLKLYP 59
DB 961 MYALPEKEVVTATRVNNGYGETKGAIIHSLXAATRTFGNDYQKGIGGAFDLKQLKLYP 1020
QY 60 EIFTKKQVSTGVAIDPSOKITEWSAKYFNGTNIHRGSGYVLKADGGQYVNLGTTT-KQF 119
DB 1021 ELFAKQISTGVPMDSVKIKQWSAKYFNGTNIHRGSGYVLKADGGQYVNLGTTT-KQF 1080
QY 120 PIQLTGEKKQKNEGFEVKG--NDG-NYYFYDLAGNMVKNFTIEDSVG-NWYFFDQDGKMYE 175
DB 1081 PKSLV-NPNHGTSSSVTGLVFDGKGYVYSTSGNAKNAFI--SLGNNWYFYDNNGYMVT 1137

Query Match      31.2%; Score 445.5; DB 2; Length 1455;
Best Local Similarity 39.9%; Pred. No. 1.4e-21;
Matches 115; Conservative 32; Mismatches 96; Indels 45; Gaps 13;

QY 1 IYNLPKEVATVTRVDDRGVNVKDAIINNLLYVNV-TIGGGEYQKKYGGAFDLKQLKLYP 59
DB 961 MYALPEKEVVTATRVNNGYGETKGAIIHSLXAATRTFGNDYQKGIGGAFDLKQLKLYP 1020
QY 60 EIFTKKQVSTGVAIDPSOKITEWSAKYFNGTNIHRGSGYVLKADGGQYVNLGTTT-KQF 119
DB 1021 ELFAKQISTGVPMDSVKIKQWSAKYFNGTNIHRGSGYVLKADGGQYVNLGTTT-KQF 1080
QY 120 PIQLTGEKKQKNEGFEVKG--NDG-NYYFYDLAGNMVKNFTIEDSVG-NWYFFDQDGKMYE 175
DB 1081 PKSLV-NPNHGTSSSVTGLVFDGKGYVYSTSGNAKNAFI--SLGNNWYFYDNNGYMVT 1137
```

```
Query Match      31.2%; Score 445.5; DB 2; Length 1390;
Best Local Similarity 39.9%; Pred. No. 1.4e-21;
Matches 115; Conservative 32; Mismatches 96; Indels 45; Gaps 13;

QY 1 IYNLPKEVATVTRVDDRGVNVKDAIINNLLYVNV-TIGGGEYQKKYGGAFDLKQLKLYP 59
DB 961 MYALPEKEVVTATRVNNGYGETKGAIIHSLXAATRTFGNDYQKGIGGAFDLKQLKLYP 1020
QY 60 EIFTKKQVSTGVAIDPSOKITEWSAKYFNGTNIHRGSGYVLKADGGQYVNLGTTT-KQF 119
DB 1021 ELFAKQISTGVPMDSVKIKQWSAKYFNGTNIHRGSGYVLKADGGQYVNLGTTT-KQF 1080
QY 120 PIQLTGEKKQKNEGFEVKG--NDG-NYYFYDLAGNMVKNFTIEDSVG-NWYFFDQDGKMYE 175
DB 1081 PKSLV-NPNHGTSSSVTGLVFDGKGYVYSTSGNAKNAFI--SLGNNWYFYDNNGYMVT 1137

Query Match      31.2%; Score 445.5; DB 2; Length 1455;
Best Local Similarity 39.9%; Pred. No. 1.4e-21;
Matches 115; Conservative 32; Mismatches 96; Indels 45; Gaps 13;

QY 1 IYNLPKEVATVTRVDDRGVNVKDAIINNLLYVNV-TIGGGEYQKKYGGAFDLKQLKLYP 59
DB 961 MYALPEKEVVTATRVNNGYGETKGAIIHSLXAATRTFGNDYQKGIGGAFDLKQLKLYP 1020
QY 60 EIFTKKQVSTGVAIDPSOKITEWSAKYFNGTNIHRGSGYVLKADGGQYVNLGTTT-KQF 119
DB 1021 ELFAKQISTGVPMDSVKIKQWSAKYFNGTNIHRGSGYVLKADGGQYVNLGTTT-KQF 1080
QY 120 PIQLTGEKKQKNEGFEVKG--NDG-NYYFYDLAGNMVKNFTIEDSVG-NWYFFDQDGKMYE 175
DB 1081 PKSLV-NPNHGTSSSVTGLVFDGKGYVYSTSGNAKNAFI--SLGNNWYFYDNNGYMVT 1137
```

```
QY 176 NKFHVDVDSYGEKGYFFFLKNGVSPRGGVLQV-----DNQTY-----Y 213
DB 1138 GAQGIN-----GANYFELNSGQLRNLAYDNGKVLSYGNDGRRYENGYYLFGQWRY 1191
QY 214 FDNYGKVRNQITNAGAMLYTLDENGKLKASVNSDAEYPTSTDVGKM 261
DB 1192 FQN-GIMAVGLTRVHGAVQY-FDASG-----FQAKGQFITAD-GKL 1230

RESULT 14
O69391 PRELIMINARY; PRT; 1455 AA.
AC O69391;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE GLUCOSYLTRANSFERASE-SI.
GN GTEC.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-WT4467;
RX MEDLINE-98231643; PubMed-9570124;
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Sobue S.,
RA Kimura S., Hamada S.;
RT "Molecular analyses of glucosyltransferase genes among strains of
RL FEMS Microbiol. Lett. 161:331-336(1998).
DR EMBL; D89978; BAA26120.1; -.
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 9.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Transferase.
SQ SEQUENCE 1455 AA; 162804 MW; 683A359D873E9E1A CRC64;

Query Match 31.2%; Score 445.5; DB 2; Length 1455;
Best Local Similarity 39.9%; Pred. No. 1.4e-21;
Matches 115; Conservative 32; Mismatches 96; Indels 45; Gaps 13;

QY 1 IYNLPGEVATVTRVDDRGVWKNDAIINNLYVN-TIGGGEYQKYGGAFLDKLQKLYP 59
DB 961 MYALPEKEVVTATRVDKYGTVPVAGSQIKNTLYVVDGSKSGKQQAQYGGAFLELQAKYP 1020
QY 60 EIFTKQVSTGVAIDPSOKITWSAKYFNGTNIHRSQVYLKADGGQYVNLGTTTKQFL 119
DB 1021 ELFPARKIISTGVPMDSVKIKQWSAKYFNGTNIHRSQVYLKADGGQYVNLGTTTKQFL 1080
QY 120 PIQLTGEKKGQNEGFYK--NDG-NYFYDLAGNMVKNFTIEDSVG-NWYFFDQDGKWE 175
DB 1081 PKSLV-NPNHGTSSSVTGLVFDGKGVYVYSTSGNQAQNAFI--SLGNWYFDNGNGYMT 1137
QY 176 NKFHVDVDSYGEKGYFFFLKNGVSPRGGVLQV-----DNQTY-----Y 213
DB 1138 GAQGIN-----GANYFELNSGQLRNLAYDNGKVLSYGNDGRRYENGYYLFGQWRY 1191
QY 214 FDNYGKVRNQITNAGAMLYTLDENGKLKASVNSDAEYPTSTDVGKM 261
DB 1192 FQN-GIMAVGLTRVHGAVQY-FDASG-----FQAKGQFITAD-GKL 1230

RESULT 15
O69397 PRELIMINARY; PRT; 1455 AA.
AC O69397;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE GLUCOSYLTRANSFERASE-SI.
```

```
GN GTEC.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-WT4467;
RX MEDLINE-98231643; PubMed-9570124;
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Sobue S.,
RA Kimura S., Hamada S.;
RT "Molecular analyses of glucosyltransferase genes among strains of
RL FEMS Microbiol. Lett. 161:331-336(1998).
DR EMBL; D89978; BAA26120.1; -.
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 9.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Transferase.
SQ SEQUENCE 1455 AA; 162914 MW; A1263427BF24B8E1 CRC64;

Query Match 31.2%; Score 445.5; DB 2; Length 1455;
Best Local Similarity 39.9%; Pred. No. 1.4e-21;
Matches 115; Conservative 32; Mismatches 96; Indels 45; Gaps 13;

QY 1 IYNLPGEVATVTRVDDRGVWKNDAIINNLYVN-TIGGGEYQKYGGAFLDKLQKLYP 59
DB 961 MYALPEKEVVTATRVDKYGTVPVAGSQIKNTLYVVDGSKSGKQQAQYGGAFLELQAKYP 1020
QY 60 EIFTKQVSTGVAIDPSOKITWSAKYFNGTNIHRSQVYLKADGGQYVNLGTTTKQFL 119
DB 1021 ELFPARKIISTGVPMDSVKIKQWSAKYFNGTNIHRSQVYLKADGGQYVNLGTTTKQFL 1080
QY 120 PIQLTGEKKGQNEGFYK--NDG-NYFYDLAGNMVKNFTIEDSVG-NWYFFDQDGKWE 175
DB 1081 PKSLV-NPNHGTSSSVTGLVFDGKGVYVYSTSGNQAQNAFI--SLGNWYFDNGNGYMT 1137
QY 176 NKFHVDVDSYGEKGYFFFLKNGVSPRGGVLQV-----DNQTY-----Y 213
DB 1138 GAQGIN-----GANYFELNSGQLRNLAYDNGKVLSYGNDGRRYENGYYLFGQWRY 1191
QY 214 FDNYGKVRNQITNAGAMLYTLDENGKLKASVNSDAEYPTSTDVGKM 261
DB 1192 FQN-GIMAVGLTRVHGAVQY-FDASG-----FQAKGQFITAD-GKL 1230

RESULT 16
Q55263 PRELIMINARY; PRT; 1590 AA.
AC Q55263;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE GTF-1.
GN GLUCOSYLTRANSFERASE.
OS Streptococcus sobrinus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OX NCBI_TaxID=1310;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 33478;
RA Sato S.;
RT "DNA and amino-acid sequences of water-insoluble-glucan synthetase
RT produced from Streptococcus sobrinus ATCC 33478.";
RL Ann. Kagoshima Univ. Dental School 16:23-29(1996).
DR EMBL; D63570; BAA09792.1; -.
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 15.
DR Pfam; PF02324; Glyco_hydro_70; 1.
```



```

RC STRAIN-WT4239;
RX MEDLINE=98231643; PubMed=9570124;
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,
RT Kimura S., Hanada S.;
RM "Molecular analyses of glucosyltransferase genes among strains of
RT Streptococcus mutans.";
RL FEMS Microbiol. Lett. 161:331-336(1998).
DR FEMS Microbiol. Lett. 161:331-336(1998).
DR ENBL; D88658; BAA26110.1; -.
DR InterPro; IPR002479; CW_binding.
DR DR InterPro; IPR003318; Glyco_hydro_70.
DR pfam; PF01473; CW_binding_1; 10.
DR pfam; PF02324; Glyco_hydro_70; 1.
DR transferase.
KW KW
SQ SEQUENCE 1455 AA; 163046 MW; 6D90A4978D35DD82 CRC64;

Query Match      30.7%; Score 438.5; DB 2; Length 1455;
Best Local Similarity 39.6%; Pred. No. 4.2e-21;
Matches 114; Conservative 32; Mismatches 97; Indels 45; Gaps 13;

QY 1 IYNLPGEVATVTVDGRGNWKKDAIINHHLYVN-TIGGGEYOKKYGGAFLDKLQLYP 59
Db :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DY 961 MYAPPEKEWTATRVDRKYTPVAGSQIKNTLYVVGSKSGKQQAKYGGAFLEELQAKYP 1020
QY 60 EIFPKKQVSTGVAIDPSQKITSESAKFNGFNILHRSGSYVLKADGGYYNLGTTKQFL 119
Db :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DY 1021 ELFARKQISTGVPMNDPSVKIKQWSAKFYENGFTNLRGAGYVLKDQATWTFSLYSNDNFL 1080
QY 120 PIQTUGBKKOGNEGFVG--NDG-NYIFYDLAGNMVKNFTIEDSVG-NWYFFDQDGKMYE 175
Db :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DY 1081 PKSLV-PPNHGTSSTVGLVFDPKGXYVYSTGSYQAKNTFI--SLGNMWYFDNNGYMT 1137
QY 176 NKHPVDVDSYEKGTYPFLKNGVSFRGLVQT-----DNQYF-----Y 213
Db ::||| |||::---GANVYFLSNGIQLRNAIDNGNKVLSYNGDGRYENGYLFGQWRX 1191
QY 214 PDNYGKMVRNOTINAGAMITYTLDENGKLIIKASNSDAEYPTSTDVGKM 261
Db :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DY 1192 PQN-GIMAVGLTRHGVQY-FDASG-----FOAQGOFITTD-GKL 1230

RESULT 20
Q00599 PRELIMINARY; PRT; 1599 AA.
ID Q00599 AC
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DI 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE GLUCOSYLTRANSFERASE S PRECURSOR (EC 2.4.1.5) (GTF) (DEXTRANSUCRASE)
DE DE (SUCCROSE 6-GLUCOSYLTRANSFERASE).
OS GN GTFK.
OS Streptococcus salivarius.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1304;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-ATCC 25975;
RX MEDLINE=93381257; PubMed=8371114;
RA Giffard P.M., Allen D.M., Milward C.P., Simpson C.L., Jacques N.A.;
RT "Sequence of the gtfK gene of Streptococcus salivarius ATCC 25975 and evolution of the gtf genes of oral streptococci.";
J. Gen. Microbiol. 139:1511-1522(1993).
[2]
RP SEQUENCE FROM N.A.
RX STRAIN-ATCC 25975;
RX MEDLINE=92148377; PubMed=1838391;
RA Giffard P.M., Simpson C.L., Milward C.P., Jacques N.A.;
RT "Molecular characterization of a cluster of at least two glucosyltransferase genes in Streptococcus salivarius ATCC 25975.";
J. Gen. Microbiol. 137:2577-2593(1991).
CC CC -!- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE OF
```

```

CC      THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE
CC      AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.
CC      -I- CATALYTIC ACTIVITY: SUCROSE + ((1,6)-ALPHA-D-GLUCOSYL)(N) - D-
CC      FRUCTOSE + ((1,6)-ALPHA-D-GLUCOSYL)(N+1).
CC      -I- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC      -I- DISEASE: DENTAL CARIES.
CC      -I- SIMILARITY: TO REGIONS OF BARLEY AND BACILLUS AMYLOLIQUEFACIENS
CC      ALPHA AMYLASES AND RABBIT GLYCOGEN PHOSPHORYLASE.
DR      EMBL; Z11872; CAA77898.1; -.
DR      EMBL; Z11873; CAA77901.1; -.
DR      EMBL; M64111; AAA26897.1; -.
DR      InterPro; IPR002479; CW_binding.
DR      Pfam; PF01473; CW_binding_1; 14.
DR      Pfam; PF02324; Glyco_hydro_70; 1.
DR      Transferrase; Glycosyltransferase; Signal; Repeat; Dental Caries.
RW      SIGNAL          1    42    POTENTIAL.
FT      CHAIN           43   1599    GLYCOSYLTRANSFERASE S.
FT      SIMILAR         867   967    TO ALPHA-AMYLASES.
FT      SWIMLAR        1132  1219    TO GLYCOGEN PHOSPHORYLASE.
FT      SEQUENCE       1599 AA; 176480 MW; 24B77869E152B707 CRC64;
SQ

Query Match              30.6%; Score 437; DB 2; Length 1599;
Best Local Similarity 43.0%; Pred. No. 5.9e-21;
Matches 105; Conservative 28; Mismatches 91; Indels 20; Gaps

QY      1 IYMLPGKEVATVRVDNRGNVKKDALINNLVVNT-IIGCGYKGYKGGAFLDKLQKLXP 59
DB      965 LYTLPGKEVVTATRTDTHGKVLDLDTSLVNKLIVYTKSSGNDFAQYGGAFDLKQLXP 1024
QY      60 EIFTKOVSVGVAIDPSQKITEMSAKFNGTNIHLRGSYVLKADGGOYNM---GTTTK 116
DB      1025 EIFKEVMASGKTIDPSVIKOWEAKYFGTGTIQKSGDYVL-SDGKLYFTVNDKGT-- 1080
QY      117 QFLPTLTITGKKOGCSGFVKGNUGNYFYFDLAGNMVKTFIEDSGWNYPFDQGGKWVEN 176
DB      1081 -FLPRAUTGTR-KATGFAVDVGVTY-VTTSQTQAQSOFVYIN-GKOYYFNQKGYLVTG 1136
QY      177 KHFVDVDSYGEKTYFFELKNKGVSRFGLQYTDONGTYFFDNYGKMYRNQTNAGAMIYTL 236
DB      1137 EQTID-----GSNYFFELPNGVMTDGVKNAKGQSLV--YGKSGLTQTGWKEVTVKD 1188
QY      237 ENKG 240
DB      1189 DSGK 1192

RESULT 21
Q00600 PRELIMINARY; PRT; 1518 AA.
ID Q00600
AC Q00600;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DI 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE GLYCOSYLTRANSFERASE I (EC 2.4.1.5) (GTF) (DEXTRANSUCRASE).
DE (SUCROSE 6-GLUCOSYLTRANSFERASE).
GN GTFU.
OS Streptococcus salivarius.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1304;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 25975;
RX MEDLINE=921148377; PubMed=1838391;
RA Giffard P.M., Simpson C.L., Milward C.P., Jacques N.A.;
RT "Molecular characterization of a cluster of at least two
RT glucosyltransferase genes in Streptococcus salivarius ATCC 25975." ;
RL J. Gen. Microbiol. 137:2377-2593(1991).
CC      -I- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT TO
CC      PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE OF
CC      THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE

```

CC AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.
CC -!- CATALYTIC ACTIVITY: SUCROSE + ((1,6)-ALPHA-D-GLUCOSYL)(N) - D-
CC FRUCTOSE + ((1,6)-ALPHA-D-GLUCOSYL)(N+1).
CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC -!- DISEASE: DENTAL CARIES.
DR EMBL: X11873; CAAT7900.1; -.
DR EMBL: M64111; AAA26896.1; -.
DR InterPro: IPR002479; CW_binding.
DR InterPro: IPR003318; Glyco_hydro_70.
DR Pfam: PF01473; CW_binding_1; 13.
DR Pfam: PF02324; Glyco_hydro_70; 1.
KW Transferase; Glycosyltransferase; Repeat; Dental caries.
FT DOMAIN 1307 1482 6 DIRECT REPEATS.
FT REPEAT 1307 1338 REPEAT 1.
FT REPEAT 1339 1352 REPEAT 2.
FT REPEAT 1372 1403 REPEAT 3.
FT REPEAT 1404 1417 REPEAT 4.
FT REPEAT 1437 1458 REPEAT 5.
FT REPEAT 1469 1482 REPEAT 6.
SQ SEQUENCE 1518 AA; 167730 MW; DAA41F717098B59A CRC64;

Query Match 30.0%; Score 428.5; DB 2; Length 1518;
Best Local Similarity 34.3%; Pred. No. 2e-20;
Matches 109; Conservative 42; Mismatches 80; Indels 87; Gaps 12;

QY 1 IYNLPGEVATVTRDGRNWKADAIINNLYVNTIGG-EYQKYGGAFDLKQKLYP 59
Db IYQLPGKEVATVTRDGRNWKADAIIDHSIYVANSKSKDYQAKYGGFEFLAEKAKYP 1064

QY 60 EITTKQVSTGVAIDPSQITKTSKAYFNGTNIHSGSGVYLKADG-GQYNNLGTTKO- 117
Db EHFKNMISTGKPIDDSYKLRQKQAKYFNGTNIHSGSGVYLKADG-GQYNNLGTTKO- 1121

QY 118 -FLPIQLTGKGGK-----QCNGKGFVKGNDGNTFYFDLAGNMV---- 152
Db NFPIQLTGKGGKVTGFGSDGKGTTFGTSGTQAKSAFVTN-GNTYFDPARGHVMVTNSE 1180

QY 153 -----KNTFTSDVGNWYFF-----DQDG---K 172
Db YSPNGKDVYRFLPFGIMLSAFYIDANGNTLYNSKRGQMYKGYTKFVDSVETDKDGKESK 1240

QY 173 MVENKHFVD-----VDSYGERGTGYFFELKNGVSFRGLGYQTDNGTYFDNY-GKMV 221
Db VVKRFYETNGVMAKGVTVIDGF-----TQYFGEDGFGQAKKLTFRGKTYFFDAHTNGGI 1296

QY 222 RNQPINAGAMIYTLDENG 239
Db KDTWRNNGKWIYFDANG 1314

RESULT 22
Q55265 PRELIMINARY; PRT; 1577 AA.
ID Q55265
AC Q55265
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE GLOCOGYLTRANSFERASE PRECURSOR.
GN GTFM.
OS Streptococcus salivarius.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1304;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95122197; PubMed=7822030;
RA Simpson C.L., Giffard P.M., Jacques N.A.;
RT "Streptococcus salivarius ATCC 25975 possesses at least two genes
RL coding for primer-independent glucosyltransferases.";
DR EMBL: L35928; AAC41413.1; -.
DR InterPro: IPR002479; CW_binding.

DR InterPro: IPR003318; Glyco_hydro_70.
DR Pfam: PF01473; CW_binding_1; 11.
DR Pfam: PF02324; Glyco_hydro_70; 1.
KW Signal; Transferase.
FT SIGNAL 1 38 POTENTIAL.
FT CHAIN 39 1577 GLUCOSYLTRANSFERASE.
FT SEQUENCE 1577 AA; 175290 MW; 3EFE898A7D3A7BF3 CRC64;

Query Match 29.2%; Score 418; DB 2; Length 1577;
Best Local Similarity 39.0%; Pred. No. 1.e-19;
Matches 105; Conservative 38; Mismatches 82; Indels 44; Gaps 11;

QY 1 IYNLPGEVATVTRDGRNWKADAIINNLYVNTIGG-EYQKYGGAFDLKQKLYP 59
Db IYQLPGKEVATVTRDGRNWKADAIINNLYVNTIGG-EYQKYGGAFDLKQKLYP 1101

QY 60 EITTKQVSTGVAIDPSQITKTSKAYFNGTNIHSGSGVYLKADG-GQYNNLGTTKO- 116
Db AIFERVOISNGRKLITNEKITQWSAKYFNGSIQGTGARYVLQDNATQYFSVRAGQT-- 1159

QY 117 QFLPIQLTGKGGKGGVFK-GNDGNYFYFDLAGNMVNTFTSDVGNWYFDQDGKHYE 175
Db -FLPKQMT---EITSGFRRVDDVQY--LSIGGYLAKNTFIQVGANQWYFDKNGNMT 1213

QY 176 NKHFVDVDSYGEKGYFFELKNGVSFRGL-----VQTDNGTY----- 212
Db GQVID-----GKK--YFFLDNGLQLRHLVRQSGDHVYYPKGVQAFNGFYDFAGPRQD 1267

QY 213 --YFNYGKMNQNTINAGAMIYTLDENG 239
Db VRYFDGNGQMYRGLHDMYCTTFYFDEKGTG 1296

RESULT 23
Q97K42 PRELIMINARY; PRT; 2817 AA.
ID Q97K42
AC Q97K42
DT 01-OCT-2001 (TREMBLrel. 18, Created)
DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE UNCHARACTERIZED PROTEIN, RELATED TO ENTEROTOXINS OF OTHER
DE CLOSTRIDIUM.
GN CAC1079.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1488;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=21359325; PubMed=11466286; Makarova K.S., Zeng Q.,
RA Noelling J., Berton G., Omerchenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.;
RT "Genome sequence and comparative analysis of the solvent-producing
RT bacterium Clostridium acetobutylicum.";
RL J. Bacteriol. 183:4823-4838(2001).
DR EMBL: AE007623; AAK79053.1; -.
DR InterPro: IPR002479; CW_binding.
DR Pfam: PF01473; CW_binding_1; 119.
KW Complete proteome.
SQ SEQUENCE 2817 AA; 318224 MW; 1851D0D4FFBEE921 CRC64;

Query Match 11.8%; Score 169; DB 16; Length 2817;
Best Local Similarity 24.9%; Pred. No. 0.0082;
Matches 77; Conservative 34; Mismatches 100; Indels 98; Gaps 18;

QY 27 INNLYVYN-----TIGGEY-----OKKYG-----GAPLQKLOK 56
Db INNNYVFNNDGVMTGWTTCNDTSKYFFDNNGVMQGLVHNNKTYFGNGDKLGLQNL 402

QY 57 L-YPEIFTKKQVSTGVAIDPSQKTEW-----SAKYENGTNHLHRSYGLKADGGY 109
 Db 403 INNYTYFD-----SGV-----MOTDWTIDSGSKYFSVNGVMTGIYI-----SGYI 448
 QY 110 NLGTTTKQFLPIQ-----LAGEKKQ-----GNEGFKGN-----DGN 142
 Db 449 GFANDKLLTGLQVINGNSYFDTNCIRLVSRWITDGDYFQDGLTDMWINDGKY 508
 QY 143 YFVLDAGNMYKNTFIEDSGVNYFFDDQDKVYENKHFVDVSYGKGYFFLKNGVSRG 202
 Db 509 YFY-IGS--VKOTGLQNDYNYFDDSGIMOTGLQKID---GK--TYFFDNGIR-QI 558
 QY 203 GLVQDTNGTYFENYCKVNRNTI-----NAGAMIYTLDENGKLIKA---SYNSDAEY 252
 Db 559 GWTYQNNKYFNSGSMQTDLKIYSYSTSPINYHYQYFGFNDGKLLGLQTIKNTYI 618
 QY 253 PTSTDVGKM 261
 Db 619 FDSNGISQM 627
 RESULT 24
 Q38658 PRELIMINARY; PRT; 316 AA.
 AC Q38658;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE E3L AMIDASE.
 GN E3L.
 OS Bacteriophage EJ-1.
 OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae.
 OX NCBI_TaxID=12402;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92380924; PubMed=1355083;
 RA Diaz E., Lopez R., Garcia J.L.;
 RT "EJ-1, a temperate bacteriophage of Streptococcus pneumoniae with a
 Myoviridae morphotype.";
 RL J. Bacteriol. 174:5516-5525(1992).
 DR EMBL; S43512; AAB23083.1; -;
 DR InterPro; IPR002479; Amidase 2.
 DR Pfam; PF01510; Amidase 2; 1.
 DR Pfam; PF01473; CW_binding_1; 5.
 SQ SEQUENCE 316 AA; 36575 MW; 42FCD64A4667ED05 CRC64;

Query Match 10.9%; Score 156; DB 9; Length 316;
 Best Local Similarity 23.5%; Pred. No. 0.0049;
 Matches 66; Conservative 46; Mismatches 93; Indels 76; Gaps 14;
 QY 8 EVAVTVTRVDRGNWKAIDAIINNLYVNTTGG-----EYQKYYG 48
 Db 48 ELGFFSHVGVNGRMVQVNVNWSW---DVGGWNAETAAVELIESHSEEFMEDY-R 103
 QY 49 AFLDKLQKLYPEITTKQVST---GVA-----IDPSQKITEN--SAKY 86
 Db 104 LYIELLRNLADEADLPKLTDDTLADGATHECTNPNPNNSHVDYFPLAKWGVSRQ 163
 QY 87 F-----NGTNIHLRSGYVLKADGGYVNLCTTKQFLPIQLTGKQKQNGFVKGN 141
 Db 164 FKQDIENGLTI---EAGW-KKNDGTWY-----VYSDGSYPKDKFEKVN-GT 205
 QY 142 YFVLDAGNMYKNTFIEDSGVNYFFDDQDKVYENKHFVDVSYGKGYFFLKNGVSR 201
 Db 206 WYFEDGSGYMLADRWKKSNDGNWYFDDQSGEMATGWKKI-----AEKMYFVDEG--AMK 258
 QY 202 GGLVQDTNGTYFENYCKVNRNTI-----NAGAMIYTLDENGKLIKA---SYNSDAEY 252
 Db 259 TGWVKYKDTWTYILDSKDGNMVSNBFIIRAGQGWYIKPDGTM 299

RESULT 25
 Q54447 PRELIMINARY; PRT; 563 AA.
 AC Q54447;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE GLUCAN-BINDING PROTEIN PRECURSOR (GBP).
 OS Streptococcus mutans.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1309;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90170123; PubMed=2307516;
 RA Banas J.A., Russell R.R., Ferretti J.J.;
 RT "Sequence analysis of the gene for the glucan-binding protein of
 Streptococcus mutans Ingritt.";
 RL Infect. Immun. 58:667-673(1990).
 CC -1- SIMILARITY: SOME, TO S.DOWNEI DEXTRANASE INHIBITOR.
 CC -1- SIMILARITY: CONTAINS SIX A AND FOUR C GLUCAN-BINDING REPEATS.
 DR EMBL; M30945; AAR26894.1; -;
 DR InterPro; IPR002479; CW_binding.
 DR Pfam; PF01473; CW_binding_1; 9.
 KW SIGNAL; Repeat.
 FT SIGNAL 1 27 POTENTIAL.
 FT CHAIN 28 563 GLUCAN-BINDING PROTEIN.
 FT DOMAIN 169 555 GLUCAN-BINDING (APPROXIMATE) (BY
 SIMILARITY).
 FT DOMAIN 169 555 6 X CIRCA 33 AA APPROXIMATE A REPEATS.
 FT REPEAT 169 201 1-1.
 FT REPEAT 264 296 1-2.
 FT REPEAT 349 382 1-3.
 FT REPEAT 425 457 1-4.
 FT REPEAT 504 535 1-5.
 FT REPEAT 544 555 1-6 (INCOMPLETE).
 FT DOMAIN 227 495 4 X 20 AA APPROXIMATE C REPEATS.
 FT REPEAT 227 246 2-1.
 FT REPEAT 312 331 2-2.
 FT REPEAT 388 407 2-3.
 FT REPEAT 479 495 2-4.
 SQ SEQUENCE 563 AA; 62932 MW; A0BB6DAAF42D1B5F CRC64;

Query Match 10.8%; Score 154.5; DB 2; Length 563;
 Best Local Similarity 21.2%; Pred. No. 0.012; 92; Indels 159; Gaps 18;
 Matches 78; Conservative 39; Mismatches 39;
 QY 17 DRGNWKAIDAIINN---LYVNV-----TIGGEY-----42
 Db 236 DSGELNKRDFVSSYADPLHYENIKHEGWFLGCDGKAAIGWRTIGKKYFDTNGVQVK 295
 QY 43 -----QKYYG-GAFDLKLOKLYPEFTTKQVSTGVAIDPSQKITEWSAKYFN 88
 Db 296 GKLISTDGNVNLISOKYKKSLDP-----DTGEA-----WTRRFVN 332
 QY 89 GTNHLRSGYV-----LKADGGYVNLCTTKQFLP---IQLTGK 127
 Db 333 AKYFFNFAGYVSTDFWYMGADGIGVTDQKIDGMDY-----FEPSSCIQVKG- 383
 QY 128 KQNGEFGVKGNDGNYFYDL-AGNMYKNTF-----IEDSVGNWYFFDQD----- 171
 Db 384 -----IABRDKGVYLDSDSGVYKVRFGTTPAERISTVREARFKTYFFGADGSRKDL 436
 QY 172 ---KMYENK---HFVDVDSYGEKTYFFLKNGVSRGLVQDTNGTYFENYKMYN --- 223
 Db 437 TGWQIIDGKTYFFDKDHSIKAKSEYSGQIGGVP-DGFAEIDGDDGYFFDTQGFVTRNFV 495
 QY 224 -----GTINAGAMIYTLDE--NGKLIKASVNS--DAEYFTSD 257
 Db 496 RKYDYSNIWYVYSDGKRVSGWQTDIGKRYFFSODEKTGKQINGQITIDGKEYTFDKD 555

Oy 258 VGMMLDON 265
|:::|
Db 556 SGEVINSN 563

Search completed: August 12, 2002, 08:08:31
Job time: 223 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 12, 2002, 08:05:34 ; Search time 49.05 Seconds
(without alignments)
4033.077 Million cell updates/sec

Title: US-09-995-749a-2

Perfect score: 1781

Sequence: 1 MEIKKFKLKSKQWTTA.....SDAEYPTSTVGKMLDQNK 1781

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 747574 seqs, 111073796 residues

Word size : 0

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database : A_Geneseq_032802.*

```

1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	15	0.9	1577	AAE91047	Alpha-D-glucosyltr
2	15	0.8	19	AAV43341	GTP antigenic epit
3	15	0.8	26	AAV43342	GTP antigenic epit
4	15	0.8	1592	AAE32925	Glucosyltransferase
5	13	0.7	22	AAE10672	L. mesenteroides a
6	13	0.7	26	AAV43343	GTP antigenic epit
7	13	0.7	26	AAV43345	GTP antigenic epit
8	13	0.7	2057	AAE10667	L. mesenteroides a
9	9	0.5	25	AAV43344	GTP antigenic epit
10	8	0.4	19	AAV43340	GTP antigenic epit
11	8	0.4	226	AAE14311	Arabidopsis thalia

12	8	0.4	226	21	AAE53230	Arabidopsis thalia
13	8	0.4	240	21	AAE53202	Arabidopsis thalia
14	8	0.4	269	20	AAV43915	Human protein kina
15	8	0.4	301	20	AAV34825	Chlamydia pneumoni
16	8	0.4	321	22	AAE99842	AGC protein kinase
17	8	0.4	338	22	ABE68636	Drosophila melanog
18	8	0.4	510	22	ABG22923	Novel human diago
19	8	0.4	540	22	ABG22922	Novel human diago
20	8	0.4	556	22	ABE68834	Putative P. abyssl
21	8	0.4	626	22	AAE94393	Human protein sequ
22	8	0.4	629	22	ABE57970	Drosophila melanog
23	8	0.4	696	21	AAE19564	Rat protein kinase
24	8	0.4	697	9	AAE82018	Protein kinase C-1
25	8	0.4	697	22	AAE78746	Human protein SEQ
26	8	0.4	730	22	AAE79730	Human protein SEQ
27	8	0.4	1068	21	AAE59273	Mouse huntingtin-1
28	8	0.4	1127	22	AAE95541	Human protein sequ
29	7	0.4	14	19	AAE39836	Light chain CDRI o
30	7	0.4	42	22	ABE30035	Peptide #2686 enco
31	7	0.4	42	22	ABE35207	Peptide #2713 enco
32	7	0.4	42	22	ABE20648	Protein #2647 enco
33	7	0.4	42	22	AAE56039	Human brain expres
34	7	0.4	42	22	AAE68410	Human bone marrow
35	7	0.4	42	22	AAE70411	Human bone marrow
36	7	0.4	42	22	AAE16225	Peptide #2659 enco
37	7	0.4	42	22	AAE38715	Peptide #2752 enco
38	7	0.4	42	22	AAE03955	Peptide #2637 enco
39	7	0.4	43	20	AAE91054	Flanking sequence
40	7	0.4	43	21	AAE27068	Beta-catenin flank
41	7	0.4	51	22	ABE42124	Peptide #9630 enco
42	7	0.4	51	22	AAE63008	Human brain expres
43	7	0.4	51	22	AAE75819	Human bone marrow
44	7	0.4	51	22	AAE35927	Peptide #9964 enco
45	7	0.4	59	22	ABE17824	Human nervous syst
46	7	0.4	65	20	AAE14449	Human secreted pro
47	7	0.4	77	19	AAE21240	Human bel2 proto-o
48	7	0.4	85	22	ABE67515	Drosophila melanog
49	7	0.4	100	19	AAE39892	Light chain of the
50	7	0.4	101	21	AAE40652	Human ORF416
51	7	0.4	103	21	AAE07714	Amino acid sequenc
52	7	0.4	103	22	AAE46037	Protonibacterium
53	7	0.4	104	22	AAE25246	Human protein sequ
54	7	0.4	105	22	AAE84107	Human immune/haema
55	7	0.4	108	22	ABE25095	Novel human diago
56	7	0.4	109	19	AAE39800	Variable domain of
57	7	0.4	117	22	ABE21656	Novel human diago
58	7	0.4	117	22	ABE28021	Novel human diago
59	7	0.4	120	21	AAE02326	Human secreted pro
60	7	0.4	122	22	ABE04086	Human musculoskele
61	7	0.4	122	22	ABE10521	Human cDNA SEQ ID
62	7	0.4	133	22	AAE18071	Human immunoglobul
63	7	0.4	133	21	AAE54920	Arabidopsis thalia
64	7	0.4	136	20	AAE73460	Human secreted pro
65	7	0.4	137	22	AAE50475	Protonibacterium
66	7	0.4	139	18	AAE27835	Amino acid sequenc
67	7	0.4	141	22	AAE74259	Human colon cancer
68	7	0.4	143	22	AAE84589	Human immune/haema
69	7	0.4	143	22	AAE01992	Human polypeptide
70	7	0.4	148	22	AAE03904	Human gene 7 encood
71	7	0.4	154	22	AAE84467	Amino acid sequenc
72	7	0.4	154	22	AAE83080	Human truncated CA
73	7	0.4	156	20	AAE06889	S. aureus glycerol
74	7	0.4	161	21	AAE25410	Pinus radiata cell
75	7	0.4	165	22	AAE67372	Protonibacterium
76	7	0.4	166	22	AAE16609	Human novel secret
77	7	0.4	182	21	AAE09240	Arabidopsis thalia
78	7	0.4	182	21	AAE52842	Arabidopsis thalia
79	7	0.4	183	21	AAE57588	Arabidopsis thalia
80	7	0.4	187	21	AAE57587	Arabidopsis thalia
81	7	0.4	188	21	AAE32721	Eucalyptus grandis
82	7	0.4	191	22	ABE69727	Drosophila melanog
83	7	0.4	192	21	AAE49874	Arabidopsis thalia
84	7	0.4	194	22	ABE23890	Novel human diago

85 Amino acid sequenc
 86 Arabidopsis thalia
 87 HTRM clone 2518753
 88 Human G protein-co
 89 Arabidopsis thalia
 90 Arabidopsis thalia
 91 Arabidopsis thalia
 92 Sequence of rat pr
 93 Type I rat brain p
 94 Mycobacterium tube
 95 Pea light-repressi
 96 Arabidopsis thalia
 97 Arabidopsis thalia
 98 Arabidopsis thalia
 99 Arabidopsis thalia
 100 Arabidopsis thalia

ALIGNMENTS

RESULT 1
 AAR91047
 ID AAR91047 standard; Protein; 1577 AA.

XX AAR91047;
 AC AAR91047;
 DT 22-MAY-1996 (first entry)
 DE Alpha-D-glucosyltransferase.

XX Alpha-D-glucosyltransferase; primer-independent; soluble glucan;
 KW sucrose; transgenic plant; cloning; Escherichia coli;
 KW phage lambda-cl13; vector; plasmid pGS501; plasmid pGS502;
 KW gene transfer; crop improvement; storage carbohydrate; pasture;
 KW feedstuff; senescence; dextran; binder; food; pharmaceutical.
 XX Streptococcus salivarius strain ATCC 25975.

OS WO9606173-A1.
 PN 29-FEB-1996.
 PD 24-AUG-1995; 95WO-AU00527.
 PF 24-AUG-1994; 94AU-0007643.
 PR (GIFP/) GIFFARD P M.
 PA (JACQ/) JACQUES N A.
 PA (SIMP/) SIMPSON C L.

XX Giffard PM, Jacques NA, Simpson CL;
 PI WPI; 1996-151376/15.
 DR N-PSDB; AAT13139.
 XX Plants contg. new bacterial DNA encoding glucosyl transferase
 PT activity - retain higher levels of stored carbohydrate(s) in a form
 PT readily digestible by ruminants
 XX Claim 4; Page 16-20; 31pp; English.

XX The sequence represents an alpha-D-glucosyltransferase from
 CC Streptococcus salivarius. The enzyme is primer-independent, and
 CC produces soluble glucan from sucrose. A gene encoding the enzyme
 CC may be cloned and expressed in Escherichia coli using a subclone
 CC of phage lambda-cl13, e.g. plasmid pGS501 or plasmid pGS502. The
 CC DNA may also be expressed in a transgenic plant, to improve the
 CC level of stored carbohydrate in a pasture plant which normally
 CC contains low levels, or to prevent degradation of stored carbohydrate
 CC during plant senescence. Dextran may be isolated from the plant, for
 CC use as a food binder or pharmaceutical additive. Primer independence
 CC ensures that the enzyme will be functional in plants. The glucan is

CC poorly degraded in plants but easily degraded by bacteria in the rumen
 CC of grazing livestock.
 XX Sequence 1577 AA;
 SQ Query Match 0.98; Score 16; DB 17; Length 1577;
 Best Local Similarity 100.0%; Pred. NO. 7e-07;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1508 ADWVPDQIYNLPKREV 1523
 Db 1035 adwvpdqlynlpkrev 1050
 |||||

RESULT 2
 AAY43341
 ID AAY43341 standard; peptide; 19 AA.

XX AAY43341;
 AC AAY43341;
 DT 26-JAN-2000 (first entry)
 DE GTF antigenic epitope.

XX GTF; antigenic epitope; glucosyltransferase; immunogenic epitope; teeth;
 KW immunogenic composition; streptococcal glucan; antibody response;
 KW cariogenic bacteria; mutans streptococci; colonisation; caries;
 KW dental caries; immunisation; therapy.
 XX Synthetic.

XX WO9952548-A2.
 PN 21-OCT-1999.
 PD 09-APR-1999; 95WO-US07828.
 PF 10-APR-1998; 98US-0081315.

XX (LEES/) LEES A.
 PA (TAUB/) TAUBMAN M A.
 PA (SMIT/) SMITH D J.
 XX Lees A, Taubman MA, Smith DJ;
 PI WPI; 1999-620289/53.

XX Immunogenic compositions for control of dental caries, based on
 PT Streptococcus mutans components, particularly for vaccination of
 PT infants

PS Claim 11; Page 44; 54pp; English.

XX This sequence represents a glucosyltransferase (GTF) immunogenic epitope,
 CC and can be used in the immunogenic composition of the invention. The
 CC immunogenic composition (A) comprises at least one streptococcal glucan
 CC (SG) and at least one component (I) covalently coupled, (indirectly, to
 CC SG. (A) elicits antibodies (Ab) to both SG and (I). (A), and related
 CC compositions, are used to induce an antibody response to cariogenic
 CC bacteria, collectively called "mutans streptococci" (i.e. any of
 CC Streptococcus mutans, S. rattus, S. cricetus or S. sobrinus), especially
 CC to prevent colonisation, to reduce numbers of bacteria or to reduce
 CC incidence of caries, in humans or other animals. The composition can
 CC therefore be used for the treatment or prevention of caries (particularly
 CC dental caries). Abs raised against (A) can be used: (a) to detect SG, and
 CC (b) for passive immunisation, e.g. applied directly to the surface of
 CC teeth. SG is a poorly immunogenic type 2 T-cell independent antigen but
 CC when coupled to (I) it produces a conjugate which induces significant
 CC immunity to both components, resulting in immunological memory and
 CC long-lasting antibody production against SG, in adults and children.

XX Sequence 19 AA;

QY	986	DNSNPVQAQLNWL	1000
Db	12	dnsnpvqaqlnwl	26

RESULT 4
AAR32925

AC AAR32925;
XX
DT 28-JUN-1993 (first entry)

Glucosyltransferase I.
GT-1; Streptococcus; dental; caries.

OS Streptococcus sobrinus.

JP05023188-A.
02-FEB-1993.

XX
XX

XX

PA (KATO//) KATO K.

DR WPY; 1993-079449/10.



PT sobrinus DNA sequence with a

XX

xxx

CC glucosyltransferase-I (and m

CC partially digesting with Sau3AI and fractionating on agarose gel.
CC The 3-5 kbp fragment was ligated into pUC18 and E. coli JM109
CC transformed with it. A GT-1 expressing clone was isolated and
CC

CC dental caries.

```

SQ      Sequence      1592 AA;

Query Match      0.8%; Score 15; DB 14; Length 1592;
Best Local Similarity 100.08; Pred. No. 7.8e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

DT 19-JAN-2001 (first entry)

Alternan sucrose; glucosyltransferase; fructose; cosmetic; foodstuff;


```

XX PA (LEES/) LEES A.
XX PA (TAUB/) TAUBMAN M A.
XX PA (SMIT/) SMITH D J.
XX PI Lees A, Taubman MA, Smith DJ;
XX XX WPI; 1999-620289/53.
XX XX
XX XX Immunogenic compositions for control of dental caries, based on
XX PT Streptococcus mutans components, particularly for vaccination of
XX PT infants
XX XX
XX PS Claim 11; Page 44; 54pp; English.
XX CC This sequence represents a glucosyltransferase (GTF) immunogenic epitope,
XX CC and can be used in the immunogenic composition of the invention. The
XX CC immunogenic composition (A) comprises at least one streptococcal glucan
XX CC (SG) and at least one component (I) covalently coupled, (in)directly, to
XX CC SG. (A) elicits antibodies (Ab) to both SG and (I). (A), and related
XX CC compositions, are used to induce an antibody response to cariogenic
XX CC bacteria, collectively called "mutans streptococci" (i.e. any of
XX CC Streptococcus mutans, S. rattus, S. cricetus or S. sobrinus), especially
XX CC to prevent colonisation, to reduce numbers of bacteria or to reduce
XX CC incidence of caries, in humans or other animals. The composition can
XX CC therefore be used for the treatment or prevention of caries (particularly
XX CC dental caries). Abs raised against (A) can be used: (a) to detect SG, and
XX CC (b) for passive immunisation, e.g. applied directly to the surface of
XX CC teeth. SG is a poorly immunogenic type 2 T-cell independent antigen but
XX CC when coupled to (I) it produces a conjugate which induces significant
XX CC immunity to both components, resulting in immunological memory and
XX CC long-lasting antibody production against SG, in adults and children.
XX XX
XX SQ Sequence 26 AA;

Query Match 0.7%; Score 13; DB 20; Length 26;
Best Local Similarity 100.0%; Pred. No. 2.2e-05;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 986 DNSNPVQAEQLN 998
DB 12 dnsnpvqaeqln 24
|||||

RESULT 8
AAB10667
ID AAB10667 standard; Protein: 2057 AA.
XX AC AAB10667;
XX 19-JAN-2001 (first entry)
XX XX
XX XX L. mesenteroides alternan sucrose protein.
XX DE
XX XX Alternan sucrose; glucosyltransferase; fructose; cosmetic; foodstuff;
XX KW syrup.
XX XX
XX XX Leuconostoc mesenteroides.
XX OS
XX XX DE19905069-A1.
XX PN
XX XX 10-AUG-2000.
XX PD
XX XX 08-FEB-1999; 99DE-1005069.
XX PF
XX XX 08-FEB-1999; 99DE-1005069.
XX PR
XX XX (PLAN-) PLANTEC BIOTECHNOLOGIE GMBH.
XX PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX PT Kossmann J, Welsh T, Quanz M, Kouth K;
XX PI
XX XX

```

```

DR WPI: 2000-550294/51.
DR N-PSDB; AAA97904.
XX
XX PT New nucleic acid encoding recombinant Leuconostoc mesenteroides
XX PT alternan sucrose protein and methods of alternan and fructose
XX PT production
XX XX
XX PS Claim 1a; Page 30-36; 64pp; German.
XX XX
XX CC This invention describes a novel nucleic acid molecule (I) encoding an
XX CC alternan sucrose (E.C. 2.4.1.140 - an enzyme, that belongs to the
XX CC glucosyltransferase group) The recombinant, purified alternan sucrose
XX CC gene is useful for the fermentative production of alternan (a
XX CC carbohydrate) and/or fructose by secreting the enzyme into a
XX CC saccharose-containing culture medium. Alternatively, the enzyme is
XX CC contacted with a saccharose-containing solution. The alternan and/or
XX CC fructose is then isolated from the medium. Cosmetic products or
XX CC foodstuffs containing alternan can be produced. Recombinant production of
XX CC alternan sucrose is advantageous as it provides a cost effective means of
XX CC producing fructose for high fructose containing syrups, production of
XX CC which previously has been achieved by costly production from maize
XX CC starch. This sequence represents the Leuconostoc mesenteroides alternan
XX CC sucrose protein which is described in the method of the invention.
XX XX
XX SQ Sequence 2057 AA;

Query Match 0.7%; Score 13; DB 21; Length 2057;
Best Local Similarity 100.0%; Pred. No. 0.0012;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 978 EFLANDIDNSNP 990
DB 589 eflandidnsnp 601
|||||

RESULT 9
AAY43344
ID AAY43344 standard; peptide; 25 AA.
XX AC AAY43344;
XX 26-JAN-2000 (first entry)
XX DT
XX XX
XX DE GTF antigenic epitope.
XX XX
XX KW GTF; antigenic epitope; glucosyltransferase; immunogenic epitope; teeth;
XX KW immunogenic composition; streptococcal glucan; antibody response;
XX KW cariogenic bacteria; mutans streptococci; colonisation; caries;
XX KW dental caries; immunisation; therapy.
XX XX
XX OS Synthetic.
XX XX
XX XX WC9952548-A2.
XX PN
XX XX 21-OCT-1999.
XX PD
XX XX 09-APR-1999; 99WO-US07828.
XX PF
XX XX 10-APR-1998; 98US-0081315.
XX PR
XX XX (LEES/) LEES A.
XX PA (TAUB/) TAUBMAN M A.
XX PA (SMIT/) SMITH D J.
XX XX
XX PI Lees A, Taubman MA, Smith DJ;
XX XX WPI; 1999-620289/53.
XX DR
XX XX Immunogenic compositions for control of dental caries, based on
XX PT Streptococcus mutans components, particularly for vaccination of
XX PT infants
XX PT
XX XX

```

PS Claim 11; Page 44; 54pp; English.

XX This sequence represents a glucosyltransferase (GTF) immunogenic epitope, and can be used in the immunogenic composition of the invention. The immunogenic composition (A) comprises at least one streptococcal glucan (SG) and at least one component (I) covalently coupled, (indirectly, to SG. (A) elicits antibodies (Ab) to both SG and (I). (A), and related compositions, are used to induce an antibody response to cariogenic bacteria, collectively called "mutans streptococci" (i.e. any of Streptococcus mutans, S. rattus, S. cricetus or S. sobrinus), especially to prevent colonisation, to reduce numbers of bacteria or to reduce incidence of caries, in humans or other animals. The composition can therefore be used for the treatment or prevention of caries (particularly dental caries). Abs raised against (A) can be used: (a) to detect SG, and (b) for passive immunisation, e.g. applied directly to the surface of teeth. SG is a poorly immunogenic type 2 T-cell independent antigen but when coupled to (I) it produces a conjugate which induces significant immunity to both components, resulting in immunological memory and long-lasting antibody production against SG, in adults and children.

XX SQ Sequence 25 AA;

Query Match 0.5%; Score 9; DB 20; Length 25;
Best Local Similarity 100.0%; Pred. No. 0.33;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 992 VQAEQLNLWL 1000
DB 17 vqaeqlnlw 25
|||||

RESULT 10
AAV43340
ID AAV43340 standard; peptide; 19 AA.

XX AC AAV43340;

XX DT 26-JAN-2000 (first entry)

XX DE GTF antigenic epitope.

XX GTF; antigenic epitope; glucosyltransferase; immunogenic epitope; teeth;
immunogenic composition; streptococcal glucan; antibody response;
cariogenic bacteria; mutans streptococci; colonisation; caries;
dental caries; immunisation; therapy.

XX OS Synthetic.

XX PN WO952548-A2.

XX PD 21-OCT-1999.

XX PF 09-APR-1999; 99WO-US07828.

XX PR 10-APR-1998; 98US-0081315.

XX PA (LEES/) LEES A.

XX PA (TAUB/) TAUBMAN M. A.

XX PA (SMIT/) SMITH D. J.

XX PI Lees A, Taubman MA, Smith DJ;

XX PD WPI; 1999-620289/53.

XX Immunogenic compositions for control of dental caries, based on Streptococcus mutans components, particularly for vaccination of infants -

XX SQ Claim 11; Page 44; 54pp; English.

XX This sequence represents a glucosyltransferase (GTF) immunogenic epitope, and can be used in the immunogenic composition of the invention. The

CC immunogenic composition (A) comprises at least one streptococcal glucan (SG) and at least one component (I) covalently coupled, (indirectly, to SG. (A) elicits antibodies (Ab) to both SG and (I). (A), and related compositions, are used to induce an antibody response to cariogenic bacteria, collectively called "mutans streptococci" (i.e. any of Streptococcus mutans, S. rattus, S. cricetus or S. sobrinus), especially to prevent colonisation, to reduce numbers of bacteria or to reduce incidence of caries, in humans or other animals. The composition can therefore be used for the treatment or prevention of caries (particularly dental caries). Abs raised against (A) can be used: (a) to detect SG, and (b) for passive immunisation, e.g. applied directly to the surface of teeth. SG is a poorly immunogenic type 2 T-cell independent antigen but when coupled to (I) it produces a conjugate which induces significant immunity to both components, resulting in immunological memory and long-lasting antibody production against SG, in adults and children.

XX SQ Sequence 19 AA;

Query Match 0.43; Score 8; DB 20; Length 19;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 986 DNSNPVVQ 993
DB 12 dnsnpvq 19
|||||

RESULT 11
AAG14311
ID AAG14311 standard; Protein; 226 AA.

XX AC AAG14311;

XX DT 17-OCT-2000 (first entry)

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 14124.

XX KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

XX OS Arabidopsis thaliana.

XX PN EP1033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-0301439.

XX PR 25-FEB-1999; 99US-0121825.

XX PR 05-MAR-1999; 98US-0123180.

XX PR 09-MAR-1999; 99US-0123548.

XX PR 23-MAR-1999; 99US-0125788.

XX PR 25-MAR-1999; 99US-0126264.

XX PR 29-MAR-1999; 99US-0126785.

XX PR 01-APR-1999; 99US-0127462.

XX PR 06-APR-1999; 99US-0128234.

XX PR 08-APR-1999; 99US-0128714.

XX PR 16-APR-1999; 99US-0129845.

XX PR 19-APR-1999; 99US-0130077.

XX PR 21-APR-1999; 99US-0130449.

XX PR 23-APR-1999; 99US-0130510.

XX PR 28-APR-1999; 99US-0130891.

XX PR 30-APR-1999; 99US-0132048.

XX PR 04-MAY-1999; 99US-0132484.

XX PR 05-MAY-1999; 99US-0132485.

XX PR 06-MAY-1999; 99US-0132486.

XX PR 07-MAY-1999; 99US-0132487.

XX PR 11-MAY-1999; 99US-0132863.

XX PR 11-MAY-1999; 99US-0134256.

PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135253.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0134218.
PR 26-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.

```
PR 22-OCT-1999; 99US-0160989.
PR 23-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 26-OCT-1999; 99US-0161362.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Watch 0.4%; Score 8; DB 21; Length 226;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 957 SDYRLIGR 964
DB 21 SDYRLIGR 28

RESULT 12
AAG53230
ID AAG53230 standard; Protein; 226 AA.
XX
AC AAG53230;
XX
XX 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 67750.
XX
DE Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
XX Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0129845.
XX 19-APR-1999; 99US-0130077.
XX 21-APR-1999; 99US-0130449.
XX 23-APR-1999; 99US-0130510.
XX 23-APR-1999; 99US-0130891.
XX 28-APR-1999; 99US-0131449.
XX 30-APR-1999; 99US-0132048.
XX 04-MAY-1999; 99US-0132407.
XX 04-MAY-1999; 99US-0132484.
XX 05-MAY-1999; 99US-0132485.
XX 06-MAY-1999; 99US-0132486.
XX 06-MAY-1999; 99US-0132487.
XX 07-MAY-1999; 99US-0132863.
XX 11-MAY-1999; 99US-0134256.
XX 14-MAY-1999; 99US-0134218.
XX 14-MAY-1999; 99US-0134219.
XX 14-MAY-1999; 99US-0134221.
XX 14-MAY-1999; 99US-0134370.
XX 18-MAY-1999; 99US-0134768.
XX 19-MAY-1999; 99US-0134941.
XX 20-MAY-1999; 99US-0135124.
XX 21-MAY-1999; 99US-0135353.
XX 24-MAY-1999; 99US-0135629.
XX 25-MAY-1999; 99US-0136021.
XX 27-MAY-1999; 99US-0136392.
XX 28-MAY-1999; 99US-0136782.
XX 01-JUN-1999; 99US-0137222.
XX 03-JUN-1999; 99US-0137528.
XX 04-JUN-1999; 99US-0137502.
XX 07-JUN-1999; 99US-0137724.
XX 08-JUN-1999; 99US-0138094.
XX 10-JUN-1999; 99US-0138540.
XX 10-JUN-1999; 99US-0138847.
XX 14-JUN-1999; 99US-0139119.
XX 16-JUN-1999; 99US-0139452.
XX 16-JUN-1999; 99US-0139453.
XX 17-JUN-1999; 99US-0139492.
XX 18-JUN-1999; 99US-0139454.
XX 18-JUN-1999; 99US-0139455.
XX 18-JUN-1999; 99US-0139456.
XX 18-JUN-1999; 99US-0139457.
XX 18-JUN-1999; 99US-0139458.
XX 18-JUN-1999; 99US-0139459.
XX 18-JUN-1999; 99US-0139460.
XX 18-JUN-1999; 99US-0139461.
XX 18-JUN-1999; 99US-0139462.
XX 18-JUN-1999; 99US-0139463.
XX 18-JUN-1999; 99US-0139750.
XX 18-JUN-1999; 99US-0139763.
XX 21-JUN-1999; 99US-0139817.
XX 22-JUN-1999; 99US-0139899.
XX 23-JUN-1999; 99US-0140353.
XX 23-JUN-1999; 99US-0140354.
XX 24-JUN-1999; 99US-0140695.
XX 28-JUN-1999; 99US-0140823.
XX 29-JUN-1999; 99US-0140991.
XX 30-JUN-1999; 99US-0141287.
XX 01-JUL-1999; 99US-0141842.
XX 01-JUL-1999; 99US-0142154.
XX 02-JUL-1999; 99US-0142055.
XX 06-JUL-1999; 99US-0142390.
XX 08-JUL-1999; 99US-0142803.
XX 09-JUL-1999; 99US-0142920.
XX 12-JUL-1999; 99US-0142977.
XX 13-JUL-1999; 99US-0143542.
XX 14-JUL-1999; 99US-0143624.
XX 15-JUL-1999; 99US-0144005.
XX 16-JUL-1999; 99US-0144085.
XX 16-JUL-1999; 99US-0144086.
XX 19-JUL-1999; 99US-0144325.
XX 19-JUL-1999; 99US-0144331.
XX 19-JUL-1999; 99US-0144332.
XX 19-JUL-1999; 99US-0144333.
XX 19-JUL-1999; 99US-0144334.
XX 19-JUL-1999; 99US-0144335.
XX 20-JUL-1999; 99US-0144352.
XX 20-JUL-1999; 99US-0144632.
XX 20-JUL-1999; 99US-0144884.
XX 21-JUL-1999; 99US-0144814.
XX 21-JUL-1999; 99US-0145086.
XX 21-JUL-1999; 99US-0145088.
XX 22-JUL-1999; 99US-0145085.
XX 22-JUL-1999; 99US-0145087.
XX 22-JUL-1999; 99US-0145089.
XX 22-JUL-1999; 99US-0145192.
XX 23-JUL-1999; 99US-0145145.
XX 23-JUL-1999; 99US-0145218.
XX 23-JUL-1999; 99US-0145224.
XX 26-JUL-1999; 99US-0145276.
XX 27-JUL-1999; 99US-0145913.
XX 27-JUL-1999; 99US-0145918.
XX 27-JUL-1999; 99US-0145919.
XX 28-JUL-1999; 99US-0145951.
XX 02-AUG-1999; 99US-0146386.
```



```
PR 02-AUG-1999; 99US-0145388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.

PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 0.4%; Score 8; DB 21; Length 226;
Best Local Similarity 100.0%; Pred. No. 28; Mismatches 0; Indels 0; Gaps 0;
Matches 8; Conservative 0;

QY 957 SDYRLGR 964
Db 21 sdvrlgr 28
|||||

RESULT 13
AAG53202
ID AAG53202 standard; Protein; 240 AA.
XX
AC AAG53202;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana Protein fragment SEQ ID NO: 67712.
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 23-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 18-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
```

PR 04-JUN-1999;	99US-0137502.	PR 06-AUG-1999;	99US-0147303.
PR 07-JUN-1999;	99US-0137724.	PR 06-AUG-1999;	99US-0147416.
PR 08-JUN-1999;	99US-0138094.	PR 09-AUG-1999;	99US-0147493.
PR 10-JUN-1999;	99US-0138540.	PR 09-AUG-1999;	99US-0147935.
PR 10-JUN-1999;	99US-0138847.	PR 10-AUG-1999;	99US-0148171.
PR 14-JUN-1999;	99US-0139119.	PR 11-AUG-1999;	99US-0148319.
PR 16-JUN-1999;	99US-0139452.	PR 12-AUG-1999;	99US-0148341.
PR 16-JUN-1999;	99US-0139453.	PR 13-AUG-1999;	99US-0148365.
PR 17-JUN-1999;	99US-0139492.	PR 13-AUG-1999;	99US-0148684.
PR 17-JUN-1999;	99US-0139454.	PR 16-AUG-1999;	99US-0149368.
PR 18-JUN-1999;	99US-0139455.	PR 17-AUG-1999;	99US-0149175.
PR 18-JUN-1999;	99US-0139456.	PR 18-AUG-1999;	99US-0149426.
PR 18-JUN-1999;	99US-0139457.	PR 20-AUG-1999;	99US-0149722.
PR 18-JUN-1999;	99US-0139458.	PR 20-AUG-1999;	99US-0149723.
PR 18-JUN-1999;	99US-0139459.	PR 20-AUG-1999;	99US-0149929.
PR 18-JUN-1999;	99US-0139460.	PR 23-AUG-1999;	99US-0149902.
PR 18-JUN-1999;	99US-0139461.	PR 23-AUG-1999;	99US-0149930.
PR 18-JUN-1999;	99US-0139462.	PR 25-AUG-1999;	99US-0150566.
PR 18-JUN-1999;	99US-0139463.	PR 26-AUG-1999;	99US-0150884.
PR 18-JUN-1999;	99US-0139750.	PR 27-AUG-1999;	99US-0151065.
PR 18-JUN-1999;	99US-0139763.	PR 27-AUG-1999;	99US-0151066.
PR 21-JUN-1999;	99US-0139817.	PR 27-AUG-1999;	99US-0151080.
PR 22-JUN-1999;	99US-0139899.	PR 30-AUG-1999;	99US-0151303.
PR 23-JUN-1999;	99US-0140353.	PR 31-AUG-1999;	99US-0151438.
PR 24-JUN-1999;	99US-0140354.	PR 01-SEP-1999;	99US-0151930.
PR 24-JUN-1999;	99US-0140695.	PR 07-SEP-1999;	99US-0152363.
PR 28-JUN-1999;	99US-0140823.	PR 10-SEP-1999;	99US-0153070.
PR 29-JUN-1999;	99US-0140991.	PR 13-SEP-1999;	99US-0153758.
PR 30-JUN-1999;	99US-0141287.	PR 15-SEP-1999;	99US-0154018.
PR 01-JUL-1999;	99US-0141842.	PR 16-SEP-1999;	99US-0154039.
PR 01-JUL-1999;	99US-0142154.	PR 20-SEP-1999;	99US-0154779.
PR 02-JUL-1999;	99US-0142055.	PR 22-SEP-1999;	99US-0155139.
PR 06-JUL-1999;	99US-0142390.	PR 23-SEP-1999;	99US-0155486.
PR 08-JUL-1999;	99US-0142803.	PR 24-SEP-1999;	99US-0155659.
PR 09-JUL-1999;	99US-0142920.	PR 28-SEP-1999;	99US-0156458.
PR 12-JUL-1999;	99US-0142977.	PR 29-SEP-1999;	99US-0156596.
PR 13-JUL-1999;	99US-0143542.	PR 04-OCT-1999;	99US-0157117.
PR 14-JUL-1999;	99US-0143624.	PR 05-OCT-1999;	99US-0157753.
PR 15-JUL-1999;	99US-0144005.	PR 06-OCT-1999;	99US-0157865.
PR 16-JUL-1999;	99US-0144085.	PR 07-OCT-1999;	99US-0158029.
PR 16-JUL-1999;	99US-0144086.	PR 08-OCT-1999;	99US-0158232.
PR 19-JUL-1999;	99US-0144325.	PR 12-OCT-1999;	99US-0158369.
PR 19-JUL-1999;	99US-0144331.	PR 13-OCT-1999;	99US-0159293.
PR 19-JUL-1999;	99US-0144332.	PR 13-OCT-1999;	99US-0159294.
PR 19-JUL-1999;	99US-0144333.	PR 13-OCT-1999;	99US-0159295.
PR 19-JUL-1999;	99US-0144334.	PR 14-OCT-1999;	99US-0159329.
PR 19-JUL-1999;	99US-0144335.	PR 14-OCT-1999;	99US-0159330.
PR 20-JUL-1999;	99US-0144352.	PR 14-OCT-1999;	99US-0159331.
PR 20-JUL-1999;	99US-0144352.	PR 14-OCT-1999;	99US-0159637.
PR 20-JUL-1999;	99US-0144352.	PR 14-OCT-1999;	99US-0159638.
PR 21-JUL-1999;	99US-0144884.	PR 18-OCT-1999;	99US-0159584.
PR 21-JUL-1999;	99US-0144814.	PR 21-OCT-1999;	99US-0160741.
PR 21-JUL-1999;	99US-0145086.	PR 21-OCT-1999;	99US-0160767.
PR 21-JUL-1999;	99US-0145088.	PR 21-OCT-1999;	99US-0160768.
PR 22-JUL-1999;	99US-0145085.	PR 21-OCT-1999;	99US-0160770.
PR 22-JUL-1999;	99US-0145087.	PR 21-OCT-1999;	99US-0160814.
PR 22-JUL-1999;	99US-0145089.	PR 21-OCT-1999;	99US-0160815.
PR 22-JUL-1999;	99US-0145192.	PR 22-OCT-1999;	99US-0160980.
PR 23-JUL-1999;	99US-0145145.	PR 22-OCT-1999;	99US-0160981.
PR 23-JUL-1999;	99US-0145218.	PR 22-OCT-1999;	99US-0160989.
PR 23-JUL-1999;	99US-0145224.	PR 25-OCT-1999;	99US-0161404.
PR 26-JUL-1999;	99US-0145276.	PR 25-OCT-1999;	99US-0161405.
PR 27-JUL-1999;	99US-0145913.	PR 25-OCT-1999;	99US-0161406.
PR 27-JUL-1999;	99US-0145918.	PR 26-OCT-1999;	99US-0161359.
PR 27-JUL-1999;	99US-0145919.	PR 26-OCT-1999;	99US-0161360.
PR 28-JUL-1999;	99US-0145951.	PR 26-OCT-1999;	99US-0161361.
PR 02-AUG-1999;	99US-0146386.	PR 28-OCT-1999;	99US-0161920.
PR 02-AUG-1999;	99US-0146388.	PR 28-OCT-1999;	99US-0161992.
PR 02-AUG-1999;	99US-0146389.	PR 28-OCT-1999;	99US-0161993.
PR 03-AUG-1999;	99US-0147038.	PR 29-OCT-1999;	99US-0162142.
PR 04-AUG-1999;	99US-0147204.		
PR 05-AUG-1999;	99US-0147302.		
PR 05-AUG-1999;	99US-0147192.		
PR 05-AUG-1999;	99US-0147260.		

Query Match 0.4%; Score 8; DB 21; Length 240;
Best Local Similarity 100.0%; Pred. No. 29;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 957 SDYRLIGR 964
 |||||
 Db 21 sdylrlgr 28

RESULT 14
 AAY43915
 ID AAY43915 standard; Protein; 269 AA.
 AC AAY43915;
 XX AAY43915;
 DF 21-DEC-1999 (first entry)
 XX Human protein kinase #1.
 DE Prediction; secondary structure; alignment; evolutionary conservation;
 KW homology; periodicity; co-variation analysis; antigenic site;
 KW site directed mutagenesis; interaction.
 XX Homo sapiens.
 OS US5958784-A.
 PN 28-SEP-1999.
 PD 25-MAR-1992; 92US-0857224.
 PF 25-MAR-1992; 92US-0857224.
 PR (BENN/) BENNER S A.
 PA Benner SA;
 XX WPI; 1999-570766/48.
 DR Predicting the folded structure of proteins -
 PT Disclosure; Column 159-162; 113pp; English.
 PS Sequences AAY43902-Y44015 represent proteins used in a novel method of
 CC predicting the folded structure of proteins, by aligning sequences of
 CC homologous proteins and using patterns of evolutionarily conserved and
 CC varied sequences to assign positions. Positions in the alignment are
 CC assigned to the surface or inside of the folded structure, active sites,
 CC and parsing segments. Secondary structural units are assigned by
 CC identifying periodicity in the assignments, and assembled into globular
 CC form using distance constraints imposed by disulfide bridges, active
 CC site assignments and co-variation analysis. The predicted secondary
 CC structures are useful for identifying antigenic sites on a protein
 CC molecule, as guides for site directed mutagenesis studies, and for
 CC understanding the interaction of a protein with other molecules.
 XX
 SQ Sequence 269 AA;

Query Match 0.4%; Score 8; DB 20; Length 269;
 Best Local Similarity 100.0%; Pred. No. 32;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 227 GKSDVWMS 234
 |||||
 Db 185 gksvdwms 192

RESULT 15
 AAY34825
 ID AAY34825 standard; Protein; 301 AA.
 AC AAY34825;
 XX 13-SEP-1999 (first entry)

XX Chlamydia pneumoniae protein not found in C. trachomatis.
 DE Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
 XX sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
 KW vaccine; neutralising epitope.
 KW Chlamydia pneumoniae.
 OS WO9927105-A2.
 XX 03-JUN-1999.
 PD 20-NOV-1998; 98WO-IB01890.
 XX 04-NOV-1998; 98US-0107078.
 PR 21-NOV-1997; 97FR-0014673.
 PA (GEST) GENSET.
 XX Griffiths R;
 PI WPI; 1999-357842/30.
 DR Genome sequence of Chlamydia pneumoniae
 XX Page 787-788; Disclosure; 1912pp; English.
 PS AAY34584-Y35879 represent the proteins encoded by all the open reading
 CC frames in the complete genome (see AAX91990) of Chlamydia pneumoniae.
 CC C. pneumoniae causes respiratory disease such as pneumonia and
 CC bronchitis and is thought to be a contributing factor in heart
 CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema
 CC nodosum or pharyngitis. The polypeptides encoded by the open reading
 CC frames of the C. pneumoniae genome (see AAY34584-Y35879) can be used in
 CC immunogenic compositions as vaccines. Vectors containing C. pneumoniae
 CC nucleotide sequences can also be used as immunogenic compositions,
 CC especially where the vector directs the expression of a neutralising
 CC epitope of C. pneumoniae.
 XX Sequence 301 AA;
 SQ

Query Match 0.4%; Score 8; DB 20; Length 301;
 Best Local Similarity 100.0%; Pred. No. 36;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1568 LQKLYPEI 1575
 |||||
 Db 227 lqklypei 234

RESULT 16
 AAB99842
 ID AAB99842 standard; Protein; 321 AA.
 XX AAB99842;
 AC AAB99842;
 XX 20-SEP-2001 (first entry)
 DE AGC protein kinase family member PKCgamma protein sequence.
 XX Protein kinase; identification; hydrophobic pocket; interacting;
 KW cancer; diabetes; inhibition; apoptosis; tissue injury;
 KW ischaemic injury; stroke.
 XX Homo sapiens.
 OS Synthetic.
 XX WO200144497-A2.
 PN 21-JUN-2001.
 XX

PF 04-DEC-2000; 2000MO-GB04598.

XX 02-DEC-1999; 99US-0168559.

XX (UYDU-) UNIV DUNDEE.

PA Alessi D, Blondi R;

XX WPI; 2001-390252/41.

DR Identifying modulators of protein kinase (PK) activity, useful in

XX developing drugs for treating cancer or diabetes, by measuring the

PT ability of the compound to modulate or mimic the interaction of PK with

PT interacting polypeptides -

XX Disclosure: Fig 16; 180pp; English.

XX The present invention describes a method for identifying a compound that

XX modulates protein kinase activity. The method comprises measuring the

XX ability of the compound to inhibit, promote or mimic the interaction of

XX a hydrophobic pocket-containing protein kinase with an interacting

XX polypeptide. The interacting polypeptide interacts with the hydrophobic

XX pocket of the protein kinase and/or comprises the amino acid sequence

XX Phe/Tyr-Xaa-Xaa-Phe/Tyr (1). The method is useful in screening assays

XX for developing pharmaceutical compounds or drugs. Compounds, polypeptides

XX or polynucleotides from the present invention are useful in medicine,

XX particularly in the manufacture of a medicament for treating a patient

XX in need of modulation of signalling by a hydrophobic pocket-containing

XX protein kinase. Specifically, the patient has cancer or diabetes or is

XX in need of inhibition of apoptosis, e.g. a patient suffering from tissue

XX injury or ischaemic injury, including stroke. The compound or

XX composition is also useful for inhibiting the degree or rate of

XX phosphorylation by the protein kinase. The interacting polypeptide or

XX compound is useful in methods of stabilising a hydrophobic pocket-

XX containing protein kinase, where the protein kinase is exposed to the

XX compound or polypeptide. AAB99786 to AAB99947 represent amino acid

XX sequences, and AA44210 and AA44211 represent oligonucleotide sequences,

XX used in the exemplification of the present invention.

XX

SQ Sequence 321 AA;

Query Match

Best Local Similarity 0.4%; Score 8; DB 22; Length 321;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 227 GKSVDWWS 234

Db 170 gksvdwvs 177

RESULT 17

ABB68636

ID ABB68636 standard; Protein; 338 AA.

AC ABB68636;

XX 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 32700.

XX Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical.

XX Drosophila melanogaster.

OS WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

DR N-PSDB; ABL12739.

XX New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signalling and cell-cell

PT interactions -

XX Disclosure: SEQ ID NO 32700; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent

XX capable of detecting 1000 or more genes from Drosophila. The invention is

XX useful in developmental biology and in elucidating cell signalling and

XX cell-cell interactions in higher eukaryotes for the development of

XX insecticides, therapeutics and pharmaceutical drugs. The invention

XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

XX sequences (ABL01840-ABL16175) and the encoded proteins

XX (ABB57737-ABB72072).

XX The sequence data for this patent did not form part of the printed

XX specification, but was obtained in electronic format directly from WIPO

XX at ftp.wipo.int/pub/published_pat_sequences.

XX

SQ Sequence 338 AA;

Query Match

Best Local Similarity 0.4%; Score 8; DB 22; Length 338;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1284 TPNUKLGV 1291

Db 2 tpu1klgv 9

RESULT 18

ABG22923

ID ABG22923 standard; Protein; 510 AA.

AC ABG22923;

XX 18-FEB-2002 (first entry)

DE Novel human diagnostic protein #22914.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;

XX food supplement; medical imaging; diagnostic; genetic disorder.

XX Homo sapiens.

XX WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

XX 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

XX N-PSDB; AAS87110.

XX New isolated polynucleotide and encoded polypeptides, useful in

XX diagnostics, forensics, gene mapping, identification of mutations

XX responsible for genetic disorders or other traits and to assess

PT

PT biodiversity -
 XX PS Claim 20; SEQ ID NO 53282; 103pp; English.
 XX CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX CC
 SQ Sequence 510 AA;

Query Match 0.4%; Score 8; DB 22; Length 510;
 Best Local Similarity 100.0%; Pred. No. 58;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 227 GKSVDWWS 234
 II|IIII|
 Db 379 gksvdwvs 386

RESULT 19
 ID ABG22922 standard; Protein; 540 AA.
 XX AC ABG22922;
 XX DT 18-FEB-2002 (first entry)
 XX DE Novel human diagnostic protein #22913.
 XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 XX food supplement; medical imaging; diagnostic; genetic disorder.
 XX OS Homo sapiens.
 XX PN WC200175067-A2.
 XX PD 11-OCT-2001.
 XX PF 30-MAR-2001; 2001WO-US08631.
 XX PR 31-MAR-2000; 2000US-0540217.
 XX PR 23-AUG-2000; 2000US-0649167.
 XX (HYSE-) HYSEQ INC.
 XX PA Drmanac RT, Liu C, Tang YF;
 XX PI WPI; 2001-639362/73.
 XX DR N-PSDB; AAS87109.
 XX New isolated polynucleotide and encoded polypeptides, useful in
 XX diagnostics, forensics, gene mapping, identification of mutations
 XX responsible for genetic disorders or other traits and to assess
 XX biodiversity -

XX PS Claim 20; SEQ ID NO 53281; 103pp; English.
 XX CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX CC
 SQ Sequence 540 AA;

Query Match 0.4%; Score 8; DB 22; Length 540;
 Best Local Similarity 100.0%; Pred. No. 61;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 227 GKSVDWWS 234
 II|IIII|
 Db 372 gksvdwvs 379

RESULT 20
 ID AAB96834 standard; Protein; 556 AA.
 XX AC AAB96834;
 XX DT 29-OCT-2001 (first entry)
 XX DE Putative P. abyssi phenylalanyl-tRNA synthetase beta subunit.
 XX KW Hyperthermophilic archaeon; hyperthermophilic protein.
 XX OS Pyrococcus abyssi.
 XX PN FR2792651-A1.
 XX PD 27-OCT-2000.
 XX PF 21-APR-1999; 99FR-0005034.
 XX PR 21-APR-1999; 99FR-0005034.
 XX (CNRS) CNRS CENT NAT RECH SCI.
 XX (IFRE-) IFREMER INST FR RECH EXPL MER.
 XX Forterre P, Thierry JC, Prieur D, Dietrich J, Lecompte O;
 XX Querellou J, Weissenbach J, Saurin W, Heilig R;
 XX WPI; 2001-126236/14.
 XX New nucleotide sequences isolated from Pyrococcus abyssi encode
 XX proteins useful in industry -
 XX Claim 7; Pages 1628-1630; 1657pp; French.
 XX The present invention relates to the genomic sequence of Pyrococcus

CC abyss1 (see AAF86431 and AAH1223-7) and P. abyss1 proteins. P. abyss1 is
CC a hyperthermophilic archaeon, which is isolated from deep-sea
CC hydrothermal vents. The present sequence is one such P. abyss1 protein.
CC The proteins of the present invention have various potential industrial
CC uses, since the proteins are stable at very high temperatures, some up to
CC 110 degrees centigrade.
CC Note: This patent is in the same patent family as WO2000065062, which
CC contains additional sequences as shown in AAB99132-AAB99143,
CC AAH75903-AAH75920 and AAG66436.

XX Sequence 556 AA;

Query Match 0.4%; Score 8; DB 22; Length 556;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 638 VLVNGKEV 645

DB 516 vlvngkev 523

RESULT 21

AAB94393
ID AAB94393 standard; Protein; 626 AA.

XX AC AAB94393;

XX DT 26-JUN-2001 (first entry)

XX DE Human protein sequence SEQ ID NO:14957.

XX KW Human; primer: detection; diagnosis; antisense therapy; gene therapy.

XX OS Homo sapiens.

XX PN EP1074617-A2.

XX PD 07-FEB-2001.

XX PF 28-JUL-2000; 2000EP-0116126.

XX PR 29-JUL-1999; 93JP-0248036.

XX PR 27-AUG-1999; 93JP-0300253.

XX PR 11-JAN-2000; 2000JP-0118776.

XX PR 02-MAY-2000; 2000JP-0183767.

XX PR 09-JUN-2000; 2000JP-0241899.

XX PA (HELI-) HELIX RES INST.

XX PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

XX PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX DR WPI; 2001-318749/34.

XX PT Primer sets for synthesizing polynucleotides, particularly the 5602
XX full-length cDNAs defined in the specification, and for the detection
XX and/or diagnosis of the abnormality of the proteins encoded by the
XX full-length cDNAs -

XX PS Claim 8; SEQ ID 14957; 2537pp + CD ROM; English.

XX CC The present invention describes primer sets for synthesising 5602
XX full-length cDNAs defined in the specification. Where a primer set
XX comprises: (a) an oligo-dT primer and an oligonucleotide complementary
XX to the complementary strand of a polynucleotide which comprises one of
XX the 5602 nucleotide sequences defined in the specification, where the
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX of an oligonucleotide comprising a sequence complementary to the
XX complementary strand of a polynucleotide which comprises a 5'-end
XX sequence and an oligonucleotide comprising a sequence complementary to a
XX polynucleotide which comprises a 3'-end sequence, where the
XX oligonucleotide comprises at least 15 nucleotides and the combination of

CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.

XX Sequence 626 AA;

Query Match 0.4%; Score 8; DB 22; Length 626;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1435 NADFFASL 1442

DB 173 nadffasl 180

RESULT 22

ABB57970
ID ABB57970 standard; Protein; 629 AA.

XX AC ABB57970;

XX DT 26-MAR-2002 (first entry)

XX DE Drosophila melanogaster polypeptide SEQ ID NO 702.

XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical.

XX OS Drosophila melanogaster.

XX PN WO200171042-A2.

XX PD 27-SEP-2001.

XX PF 23-MAR-2001; 2001WO-US09231.

XX PR 23-MAR-2000; 2000US-191637P.

XX PR 11-JUL-2000; 2000US-0614150.

XX PA (PEKE) PE CORP NY.

XX PI Venter JC, Adams M, Li PWD, Myers EW;

XX DR WPI; 2001-656860/75.

XX DR N-PSDB; ABL02073.

XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -

XX PS Disclosure; SEQ ID NO 702; 21pp + Sequence Listing; English.

XX CC The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins
XX (ABE57737-ABE72072).

XX CC The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pat_sequences.

SQ Sequence 629 AA;

Query Match 0.4%; Score 8; DB 22; Length 629;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 TAATATVA 25
|||||||
Db 164 taatavta 171

RESULT 23

AAB19564
ID AAB19564 standard; Protein: 696 AA.

XX AC AAB19564;

XX DT 22-JAN-2001 (first entry)

XX DE Rat protein kinase C type 1 subunit.

XX KW Protein kinase C; PKC-gamma gene; neurodegenerative disorder;

XX KW Alzheimer disease; dopaminergic cell degeneration;

XX KW movement impairment; Parkinson's disease; Huntington's disease;

XX KW dementia; multiple system atrophy; progressive supranuclear palsy;

XX KW ganglionic degeneration; vascular Parkinsonism; ballism; nortropic;

XX KW neuroprotective; anticonvulsant; antiparkinsonian; gene therapy;

XX KW rat.

XX OS Rattus rattus.

XX PN W0200053218-A1.

XX PD 14-SEP-2000.

XX PF 09-MAR-2000; 2000WC-GB00860.

XX PR 09-MAR-1999; 99GB-0005218.

XX PA (UNIU) UNIV GLASGOW.

XX PI Davies RW, Payne AP, Sutcliffe RG;

XX DR WPI: 2000-594262/56.

XX DR N-PSDB: AAB88499.

Use of polypeptide comprising protein kinase C type I and polynucleotide encoding the protein for prophylaxis and treatment of neurodegenerative disorders such as Parkinson's, Alzheimer's and Huntington's diseases -

XX PS Disclosure; Fig 10; 89pp; English.

XX CC The present sequence represents the rat protein kinase C (PKC) type 1 subunit encoded by the PKC-gamma gene (see AAB88499). The invention is based on the discovery that a mutation within the PKC-gamma gene is associated with the AS/AS mutant rat. This mutant rat arose spontaneously from the parent Albino-Swiss (AS) strain and is characterized by gait disturbances. Examination of the mutated PKC-gamma gene (NNG gene), which replaced the native Glu281 codon with a stop codon, resulting in premature termination of the PKC protein at residue 280; the respiratory domain of type 1 PKC is retained, but the kinase and ATP-binding domains are not. A claimed polypeptide comprising PKC type 1 can be used to treat a degenerative disorder such as Alzheimer's disease, dopaminergic cell degeneration, movement impairment, Parkinson's disease, Huntington's disease, dementia with Lewy bodies, multiple-system atrophy, progressive supranuclear palsy, cortical-basal ganglionic (corticobasal) degeneration, vascular parkinsonism and ballism. The presence of a mutation in the PKC-gamma is used in claimed methods of testing a subject for predisposition to a neurodegenerative disorder. A limited PKC type 1 polypeptide can

CC be used to promote nervous system degeneration for the production of animal models useful e.g. for drug screening. A PKC type 1 polypeptide is used for preventing, delaying, treating or inhibiting degeneration of the nervous system, and for the identification of compounds for use in the treatment of neurodegenerative disorders.

XX SQ Sequence 696 AA;

Query Match 0.4%; Score 8; DB 21; Length 696;
Best Local Similarity 100.0%; Pred. No. 78;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 227 GKSVDMWS 234
|||||||
Db 532 gksvdmws 539

RESULT 24

AAP82018
ID AAP82018 standard; protein: 697 AA.

XX AC AAP82018;

XX DT 16-OCT-1990 (first entry)

XX DE Protein kinase C-IV (PKC-IV).

XX KW Protein kinase C; PKC; cancer; ds.

XX OS Rattus sp.

XX PN W08801303-A.

XX PD 25-FEB-1988.

XX PF 13-AUG-1987; 87WO-US02005.

XX PR 13-AUG-1986; 86US-0896476.

XX PA (GENE-) GENETICS INST INC.

XX PI Knopf JI;

XX DR WPI: 1988-064018/09.

XX DR N-NSPD; N80336.

XX PT New DNA sequences coding for protein C enzyme - and new expressed polypeptide(s), useful for detecting tumour promoting activity of test cpds.

XX PS Claim 1; Fig 2; 36pp; English.

XX CC PKC product can be used in determining tumour promoting properties of a test sample, and probes derived from the sequence may be used in isolating polypeptides from other species.

XX SQ Sequence 697 AA;

Query Match 0.4%; Score 8; DB 9; Length 697;
Best Local Similarity 100.0%; Pred. No. 78;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 227 GKSVDMWS 234
|||||||
Db 533 gksvdmws 540

RESULT 25

AAM78746

ID AAM78746 standard; Protein: 697 AA.

XX AAM78746;
XX
XX
DT 06-NOV-2001 (first entry)
XX
DE Human protein SEQ ID NO 1408.
XX
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation.
XX
OS Homo sapiens.
XX
PN W0200157190-A2.
XX
PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001WO-US04098.
XX
PR 03-FEB-2000; 2000US-0496914.
PR 27-APR-2000; 2000US-0560875.
PR 20-JUN-2000; 2000US-0598075.
PR 19-JUL-2000; 2000US-0620325.
PR 01-SEP-2000; 2000US-0651936.
PR 15-SEP-2000; 2000US-0663561.
PR 20-OCT-2000; 2000US-0693325.
PR 30-NOV-2000; 2000US-0728422.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
XX WPI: 2001-476283/51.
DR N-PSDB; AAK51879.
XX
XX Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy -
XX
XX Claim 20; Page 3672-3673; 6221pp; English.
XX
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAH80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
XX
SQ Sequence 697 AA;

Query Match 0.4%; Score 8; DB 22; Length 697;
Best Local Similarity 100.0%; Pred. No. 78;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 227 GKSDVWMS 234
DB 533 gksvdwms 540
|||||

RESULT 26
AAM79730
ID AAM79730 standard; Protein; 730 AA.

XX AAM79730;
XX
XX
DT 06-NOV-2001 (first entry)
XX
DE Human protein SEQ ID NO 3376.
XX
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation.
XX
OS Homo sapiens.
XX
PN W0200157190-A2.
XX
PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001WO-US04098.
XX
PR 03-FEB-2000; 2000US-0496914.
PR 27-APR-2000; 2000US-0560875.
PR 20-JUN-2000; 2000US-0598075.
PR 19-JUL-2000; 2000US-0620325.
PR 01-SEP-2000; 2000US-0651936.
PR 15-SEP-2000; 2000US-0663561.
PR 20-OCT-2000; 2000US-0693325.
PR 30-NOV-2000; 2000US-0728422.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
XX WPI: 2001-476283/51.
DR N-PSDB; AAK52863.
XX
XX Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy -
XX
XX Claim 20; Page 316; 6221pp; English.
XX
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAH80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
XX
SQ Sequence 730 AA;

Query Match 0.4%; Score 8; DB 22; Length 730;
Best Local Similarity 100.0%; Pred. No. 81;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 227 GKSDVWMS 234
DB 561 gksvdwms 568
|||||

RESULT 27
AAM79730
ID AAM79730 standard; Protein; 1068 AA.

XX AC AAY59273;
XX DT 17-APR-2000 (first entry)
XX DE Mouse huntingtin-interacting protein (mH1P1a).
XX KW Huntingtin Interacting Protein; HIP; death effector domain; DED; human;
XX KW apoptosis; HIP-apoptosis modulating protein; cell death; gene therapy;
XX KW Huntington's disease; nontropic; anticonvulsant; cytostatic; mouse.
XX OS Mus sp.
XX PN W09960986-A2.
XX PD 02-DEC-1999.
XX PF 27-MAY-1999; 99WO-US11743.
XX PR 27-MAY-1998; 98US-0085199.
XX PA (UYBR-) UNIV BRITISH COLUMBIA.
XX PA (MERI) MERCK FROSST CANADA INC.
XX PI Kaichman M, Hayden MR, Hackam A, Chopra V, Nicholson DW;
XX PI Vallincourt JP, Rasper DM;
XX DR WPI; 2000-097055/08.
XX DR N-PSDB; AAY58749.
XX PT Novel proteins useful for treating Huntington's disease by gene therapy
XX PT techniques, and cancers
XX PS Example 11; Page 73-77; 91pp; English.
XX CC The invention relates to Huntingtin Interacting Protein (HIP), that
XX CC includes a death effector domain (DED), suggesting apoptotic function.
XX CC Proteins with DED (referred as HIP-apoptosis modulating proteins) are
XX CC useful for inducing apoptotic death in cells. The HIP is a logical target
XX CC for therapy in Huntington's disease since it has a differential apoptotic
XX CC activity, modulated by interaction with Huntingtin having normal and
XX CC expanded repeats. HIP is also used as a therapeutic agent to introduce
XX CC apoptosis in cancer cells. Increasing expression of normal (non-expanded)
XX CC Huntingtin or the HIP-apoptotic modulator-binding portion, a modified
XX CC HIP-apoptotic modulator in which the DED has been deleted, is helpful for
XX CC treating Huntington's disease. The present sequence represents the mouse
XX CC m1P1a.
XX SQ Sequence 1068 AA;

Query Match 0.4%; Score 8; DB 21; Length 1068;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1293 DKVVLHMG 1300
| | | | |
Db 891 dkvvllmg 898

RESULT 28
AAB95541
ID AAB95541 standard; Protein; 1127 AA.
XX AC AAB95541;
XX DT 26-JUN-2001 (first entry)
XX DE Human protein sequence SEQ ID NO:18149.
XX KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX OS Homo sapiens.

XX PN EP1074617-A2.
XX PD 07-FEB-2001.
XX PF 28-JUL-2000; 2000EP-0116126.
XX PR 29-JUL-1999; 99JP-0248036.
XX PR 27-AUG-1999; 99JP-0300253.
XX PR 11-JAN-2000; 2000JP-0118776.
XX PR 02-MAY-2000; 2000JP-0183767.
XX PR 09-JUN-2000; 2000JP-0241899.
XX PA (HELI-) HELIX RES INST.
XX PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX DR WPI; 2001-318749/34.
XX PR Primer sets for synthesizing polynucleotides, particularly the 5602
XX PT full-length cDNAs defined in the specification, and for the detection
XX PT and/or diagnosis of the abnormality of the proteins encoded by the
XX PT full-length cDNAs -
XX PS Claim 8; SEQ ID 18149; 2537pp + CD ROM; English.
XX CC The present invention describes primer sets for synthesizing 5602
XX CC full-length cDNAs defined in the specification. Where a primer set
XX CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
XX CC to the complementary strand of a polynucleotide which comprises one of
XX CC the 5602 nucleotide sequences defined in the specification, where the
XX CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX CC of an oligonucleotide comprising a sequence complementary to the
XX CC complementary strand of a polynucleotide which comprises a 5'-end
XX CC sequence and an oligonucleotide comprising a sequence complementary to a
XX CC polynucleotide which comprises a 3'-end sequence, where the
XX CC oligonucleotide comprises at least 15 nucleotides and the combination of
XX CC the 5'-end sequence/3'-end sequence is selected from those defined in
XX CC the specification. The primer sets can be used in antisense therapy and
XX CC in gene therapy. The primers are useful for synthesizing polynucleotides,
XX CC particularly full-length cDNAs. The primers are also useful for the
XX CC detection and/or diagnosis of the abnormality of the proteins encoded by
XX CC the full-length cDNAs. The primers allow obtaining of the full-length
XX CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
XX CC AAH13633 to AAH18742 represent human cDNA sequences. AAB92446 to
XX CC AAB95893 represent human amino acid sequences, and AAH13629 to AAH13632
XX CC represent oligonucleotides, all of which are used in the exemplification
XX CC of the present invention.
XX SQ Sequence 1127 AA;

Query Match 0.4%; Score 8; DB 22; Length 1127;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 103 TPVSVLPS 110
| | | | |
Db 487 tpvsvlps 494

RESULT 29
AAW39836
ID AAW39836 standard; peptide; 14 AA.
XX AC AAW39836;
XX DT 16-JUN-1998 (first entry)
XX DE Light chain CDRI of catalytic antibody 8G4G.
XX KW Variable domain; lambda light chain; catalytic antibody; degradation;

KW cocaine; cocaine transistlon state analogue; TSA; benzoic acid;
KW phenyl cocaine; immunogenic conjugate; reduction; cocaine; treatment;
KW overdose; addiction.
XX
OS Mus sp.
XX WO9749800-A1.
PN
XX
PD 31-DEC-1997.
XX
XX 25-JUN-1997; 97WO-US10965.
PF
XX 25-JUN-1996; 96US-0672345.
PR
XX (UYCO) UNIV COLUMBIA NEW YORK.
PA
XX Landry DW;
PI
XX WPI; 1998-077166/07.
DR
XX New catalytic antibodies able to decompose cocaine, single-chain
PT analogues - used to treat cocaine overdose and addiction, required
PT in far smaller doses than antibodies that antagonise cocaine by
PT simply binding
XX
XX Claim 8; Page 86; 147pp; English.
PS
XX AAM39836-38 represent the sequences of the light chain complementarity
CC determining regions (CDRs) of the catalytic antibody 8G4G, which is able
CC to degrade cocaine. A series of cocaine transition state analogues
CC (TSAs) were prepared and used to immunise mice for production of
CC hybridomas. Catalytic antibodies were identified by their capacity to
CC release 3H-benzoic acid from 3H-phenyl cocaine. The 8G4G antibody was
CC identified using TSA3, and has a per minute Kcat of 0.60. The antibodies
CC reduce the concentration of cocaine in a subject, and are used
CC particularly for the treatment of an overdose. They are also used for
CC treating addiction (by reducing the in vivo concentration that can be
CC achieved).
XX
SQ Sequence 14 AA;
Query Match 0.4%; Score 7; DB 19; Length 14;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1007 GTITANN 1013
Db 5 gtitann 11
|||||
RESULT 30
ABB30035
ID ABB30035 standard; Peptide; 42 AA.
XX
XX ABB30035;
AC
XX
XX 01-FEB-2002 (first entry)
DT
XX
XX Peptide #2686 encoded by breast cell single exon nucleic acid probe.
DE
XX Human; microarray; single exon probe; gene expression; breast;
KW disease; cancer.
XX
XX Homo sapiens.
OS
XX WO200157271-A2.
PN
XX
XX 09-AUG-2001.
PD
XX
XX 30-JAN-2001; 2001WO-US00662.
PF
XX
XX 04-FEB-2000; 2000US-0180312.
PR

PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
FI
XX WPI; 2001-496933/54.
DR
XX New spatially-addressable set of single exon nucleic acid probes,
PT useful for measuring gene expression in sample derived from human
PT breast, comprises number of single exon nucleic acid probes -
XX
XX Claim 27; SEQ ID NO 13003; 327pp + sequence listing; English.
PS
XX The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human breast and BT 474 cells. The method involves contacting
CC the probes with a collection of detectably labelled nucleic acids
CC derived from mRNA of human breast, and then measuring the label
CC bound to each probe of the microarray. The probes are useful for
CC verifying the expression of regions of genomic DNA predicted to
CC encode proteins. They are useful for gene discovery, and for
CC determining predisposition and/or prognosing breast disease. Gene
CC expression analysis is useful for assessing the toxicity of chemical
CC agents on cells. The microarray of this invention presents a far greater
CC diversity of probes for measuring gene expression, with far less bias
CC than expressed sequence tag microarrays. The method is suitable for
CC rapid production of functional information from genomic sequence. The
CC present sequence is a peptide encoded by a single exon nucleic acid
CC probe of the invention.
XX Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 42 AA;
Query Match 0.4%; Score 7; DB 22; Length 42;
Best Local Similarity 100.0%; Pred. No. 66;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 19 AAVATVA 25
Db 24 aavatva 30
|||||

Search completed: August 12, 2002, 08:09:51
Job time: 257 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 12, 2002, 08:07:29 ; Search time 19.89 Seconds
(without alignments) 2187.128 Million cell updates

Title: US-09-995-749A-2

```

1 MEIKKKKFLYKSGKQWVTA .....SDAETPTSTDVGKMLDONKL 1781
perfect score: 1781
Sequence:

```

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 231628 seqs, 24425594 residues

Word size : 0

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length:	2000000000
------------------------	------------

Post-processing: Listing first 100 summaries

```
Database :
Issued Patents AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfile1.pep:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	22	1.2	1430	3	US-09-008-172-2		Sequence 2, Appl
2	22	1.2	1430	4	US-09-210-361-6		Sequence 6, Appl
3	16	0.9	1577	2	US-08-793-824-2		Sequence 2, Appl
4	15	0.8	1375	4	US-09-210-361-4		Sequence 4, Appl
5	15	0.8	1475	3	US-09-007-999-2		Sequence 2, Appl
6	15	0.8	1475	4	US-09-210-361-2		Sequence 2, Appl
7	8	0.4	269	2	US-07-857-224B-14		Sequence 14, Appl
8	8	0.4	1068	4	US-09-085-199B-11		Sequence 11, Appl
9	7	0.4	14	2	US-08-672-345C-40		Sequence 40, Appl
10	7	0.4	14	4	US-09-214-095D-40		Sequence 40, Appl
11	7	0.4	71	1	US-08-280-443-24		Sequence 24, Appl
12	7	0.4	71	1	US-08-457-459-24		Sequence 24, Appl
13	7	0.4	71	1	US-08-555-678-24		Sequence 24, Appl
14	7	0.4	71	5	PT-US95-02275-24		Sequence 24, Appl
15	7	0.4	81	2	US-08-436-420-34		Sequence 34, Appl
16	7	0.4	98	4	US-09-214-095D-4		Sequence 4, Appl
17	7	0.4	99	2	US-08-672-345C-94		Sequence 94, Appl
18	7	0.4	100	4	US-09-214-095D-96		Sequence 96, Appl
19	7	0.4	109	2	US-08-672-345C-4		Sequence 4, Appl
20	7	0.4	260	2	US-07-857-224B-7		Sequence 7, Appl
21	7	0.4	260	2	US-07-857-224B-8		Sequence 8, Appl
22	7	0.4	263	2	US-07-857-224B-13		Sequence 13, Appl
23	7	0.4	264	2	US-07-857-224B-10		Sequence 10, Appl
24	7	0.4	264	2	US-07-857-224B-11		Sequence 11, Appl
25	7	0.4	264	2	US-07-857-224B-12		Sequence 12, Appl
26	7	0.4	264	2	US-07-857-224B-15		Sequence 15, Appl
27	7	0.4	264	2	US-07-857-224B-17		Sequence 17, Appl

ALIGNMENTS

```
RESULT 1
US-09-008-172-2
; Sequence 2, Application US/09008172
; Patent No. 6127602
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starch and
; TITLE OF INVENTION: Latexes in Paper Manufacture
; FILE REFERENCE: 0357CR
; CURRENT APPLICATION NUMBER: US/09/008,172
; CURRENT FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1430
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-008-172-2

Query Match 1.2%; Score 22; DB 3; Length 1430;
Best Local Similarity 100.0%; Pred. No. 3.6e-13;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 979 FLANDIDNSNPVQAEQLNWL 1000
|
Db 420 FLANDIDNSNPVQAEQLNWL 441

RESULT 2
US-09-210-361-6
; Sequence 6, Application US/09210361
; Patent No. 6284479
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starches and
; TITLE OF INVENTION: Latexes in Paper Manufacture
; FILE REFERENCE: 0357CR
; CURRENT APPLICATION NUMBER: US/09/210,361
; CURRENT FILING DATE: 1998-12-11
; EARLIER APPLICATION NUMBER: 09/007,999
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/009,620
; EARLIER FILING DATE: 1998-01-20
; EARLIER APPLICATION NUMBER: 08/485,243
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/008,172
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 1430
; TYPE: PRT
; ORGANISM: streptococcus mutans
US-09-210-361-6

Query Match 1.2%; Score 22; DB 4; Length 1430;
Best Local Similarity 100.0%; Pred. No. 3.6e-13;
Matches, 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 979 FLANDIDNSNPVQAEQLNWL 1000
```

```
Db 420 FLANDIDNSNPVQAEQLNWL 441
|
RESULT 3
US-08-793-824-2
; Sequence 2, Application US/08793824
; Patent No. 5981838
; GENERAL INFORMATION:
; APPLICANT: Simpson, Christine Lynn
; APPLICANT: Giffard, Philip Morrison
; APPLICANT: Jacques, Nicholas Anthony
; TITLE OF INVENTION: Genetic Manipulation of Plants to
; TITLE OF INVENTION: Increase Stored Carbohydrates
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Griffith Hack & Co
; STREET: Level 8, 168 Walker Street
; CITY: No. 5981838th Sydney
; STATE: New South Wales
; COUNTRY: Australia
; ZIP: 2060
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/793,824
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: AU PM7643
; FILING DATE: 24-AUG-1994
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 61 2 9957 5944
; TELEFAX: 61 2 957 6288
; TELEX: 26547
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1577 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus salivarius
US-08-793-824-2
```

```
Query Match 0.9%; Score 16; DB 2; Length 1577;
Best Local Similarity 100.0%; Pred. No. 5e-07;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1508 ADWVPDQIYNLPQKEV 1523
|
Db 1035 ADWVPDQIYNLPQKEV 1050
```

```
RESULT 4
US-09-210-361-4
; Sequence 4, Application US/09210361
; Patent No. 6284479
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starches and
; TITLE OF INVENTION: Latexes in Paper Manufacture
; FILE REFERENCE: 0357CR
; CURRENT APPLICATION NUMBER: US/09/210,361
; CURRENT FILING DATE: 1998-12-11
; EARLIER APPLICATION NUMBER: 09/007,999
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
```

1
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/009,620
; EARLIER FILING DATE: 1998-01-20
; EARLIER APPLICATION NUMBER: 08/485,243
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/008,172
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 1375
; TYPE: PRT
; ORGANISM: streptococcus mutans
US-09-210-361-4

Query Match 0.8%; Score 15; DB 4; Length 1375;
Best Local Similarity 100.0%; Pred. No. 4.6e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 986 DNSNPVQAEQLNWL 1000
|||||

DB 439 DNSNPVQAEQLNWL 453

RESULT 5
US-09-007-999-2
; Sequence 2, Application US/09007999
; Patent No. 6087559
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starch and
; TITLE OF INVENTION: Latexes in Paper Manufacture
; FILE REFERENCE: 0356D
; CURRENT APPLICATION NUMBER: US/09/007,999
; CURRENT FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-007-999-2

Query Match 0.8%; Score 15; DB 3; Length 1475;
Best Local Similarity 100.0%; Pred. No. 4.9e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 986 DNSNPVQAEQLNWL 1000
|||||

DB 413 DNSNPVQAEQLNWL 427

RESULT 6
US-09-210-361-2
; Sequence 2, Application US/09210361
; Patent No. 6284479
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starches and
; TITLE OF INVENTION: Latexes in Paper Manufacture
; FILE REFERENCE: 0357CR
; CURRENT APPLICATION NUMBER: US/09/210,361
; CURRENT FILING DATE: 1998-12-11
; EARLIER APPLICATION NUMBER: 09/007,999
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07

; EARLIER APPLICATION NUMBER: 09/009,620
; EARLIER FILING DATE: 1998-01-20
; EARLIER APPLICATION NUMBER: 08/485,243
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/008,172
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-210-361-2

Query Match 0.8%; Score 15; DB 4; Length 1475;
Best Local Similarity 100.0%; Pred. No. 4.9e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 986 DNSNPVQAEQLNWL 1000
|||||

DB 413 DNSNPVQAEQLNWL 427

RESULT 7
US-07-857-224B-14
; Sequence 14, Application US/07857224B
; Patent No. 5958784
; GENERAL INFORMATION:
; APPLICANT: Benner, Steven A.
; TITLE OF INVENTION: Predicting Folded Structures of Proteins
; NUMBER OF SEQUENCES: 114
; CORRESPONDENCE ADDRESS:
; ADDRESSES: Steven A. Benner
; STREET: Hadlaubstrasse 151
; CITY: Zurich
; STATE: none
; COUNTRY: Switzerland
; ZIP: (note: this is an international post code) CH-8092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch diskette, 1.4 Mb storage
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 7.0
; SOFTWARE: Microsoft Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/857,224B
; FILING DATE: 03/25/92
; CLASSIFICATION: 436
; PRIOR APPLICATION DATA: none
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (International) 41 1 632 2830
; TELEFAX: (International) 41 1 262 2437
; TELEX: none
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 269
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE:
; DESCRIPTION: protein
; ORIGINAL SOURCE:
; ORGANISM: human
; FEATURE: Protein kinase; Table 8 Column 15
; PUBLICATION INFORMATION:
; AUTHORS:
; AUTHORS: Hanks, S. K.
; AUTHORS: Quinn, A. M.
; AUTHORS: Hunter, T.
; TITLE: The protein kinase family
; JOURNAL: Science
; VOLUME: 241

;
; PAGES: 42-52
; DATE: 1988
; US-07-857-224B-14

Query Match 0.4%; Score 8; DB 2; Length 269;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 227 GKSDVWWS 234
|||||||
DB 185 GKSDVWWS 192

RESULT 8
US-09-085-199B-11
; Sequence 11, Application US/09085199B
; Patent No. 6235879
; GENERAL INFORMATION:
; APPLICANT: Hayden, Michael R.
; APPLICANT: Hackam, Abigail
; APPLICANT: Hug, A.H.M. Mahbubul
; APPLICANT: Chopra, Vikramjit Singh
; APPLICANT: Kalchman, Michael
; TITLE OF INVENTION: Apoptosis Modulators That Interact with the
; TITLE OF INVENTION: Huntington's Disease Gene
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Oppedahl & Larson
; STREET: PO Box 5270
; CITY: Frisco
; STATE: CO
; COUNTRY: USA
; ZIP: 80443-5270
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Kb storage
; OPERATING SYSTEM: MS DOS 5.0
; SOFTWARE: WordPerfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/085,199B
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Larson, Marinda T.
; REGISTRATION NUMBER: 32038
; REFERENCE/DOCKET NUMBER: UBC.P-013052
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (970) 668-2050
; TELEFAX: (970) 668-2052
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1068
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: no
; ORIGINAL SOURCE:
; ORGANISM: mouse
; FEATURE:
; OTHER INFORMATION: Huntington-interacting protein -mH1pla
; US-09-085-199B-11

Query Match 0.4%; Score 8; DB 4; Length 1068;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1293 DKVVLHNG 1300
|||||||
DB 891.DKVVLHNG 898

RESULT 9
US-08-672-345C-40
; Sequence 40, Application US/08672345C
; Patent No. 5948658
; GENERAL INFORMATION:
; APPLICANT: Landry, Donald, W.
; TITLE OF INVENTION: ANTI-COCAINE CATALYTIC ANTIBODY
; NUMBER OF SEQUENCES: 108
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper and Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/672,345C
; FILING DATE: 24-JUN-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 0575/51400
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-278-0400
; TELEFAX: 212-391-0525
; INFORMATION FOR SEQ ID NO: 40:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-672-345C-40

Query Match 0.4%; Score 7; DB 2; Length 14;
Best Local Similarity 100.0%; Pred. No. 8;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1007 GTITANN 1013
|||||||
DB 5 GTITANN 11

RESULT 10
US-09-214-095D-40
; Sequence 40, Application US/09214095D
; Patent No. 6280987
; GENERAL INFORMATION:
; APPLICANT: Landry, Donald
; TITLE OF INVENTION: ANTI-COCAINE CATALYTIC ANTIBODY
; FILE REFERENCE: 51400-A-PCT-US
; CURRENT APPLICATION NUMBER: US/09/214,095D
; CURRENT FILING DATE: 1999-07-19
; NUMBER OF SEQ ID NOS: 121
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 40
; LENGTH: 14
; TYPE: PRT
; ORGANISM: Murinae gen. sp.
; US-09-214-095D-40

Query Match 0.4%; Score 7; DB 4; Length 14;
Best Local Similarity 100.0%; Pred. No. 8;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1007 GTTANN 1013
DB 5 GTTANN 11

RESULT 11
US-08-280-443-24
; Sequence 24, Application US/08280443
; Patent No. 5643778
; GENERAL INFORMATION:
; APPLICANT: Nishikura, Kazuko
; TITLE OF INVENTION: RNA Editing Enzyme and Methods of Use
; TITLE OF INVENTION: Thereof
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Cntr, P.O. Box 457
; CITY: Spring House
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/280,443
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/197,794
; FILING DATE: 17-FEB-1994
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: WST49AUSA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-540-9206
; TELEFAX: 215-540-5818
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 71 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-280-443-24

Query Match 0.4%; Score 7; DB 1; Length 71;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 283 NGKEVAR 289
DB 40 NGKEVAR 46

RESULT 12
US-08-457-459-24
; Sequence 24, Application US/08457459
; Patent No. 5677428
; GENERAL INFORMATION:
; APPLICANT: Nishikura, Kazuko
; TITLE OF INVENTION: RNA Editing Enzyme and Methods of Use
; TITLE OF INVENTION: Thereof
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Cntr, P.O. Box 457
; CITY: Spring House
; STATE: Pennsylvania
; COUNTRY: USA
```

```
QY 283 NGKEVAR 289
DB 40 NGKEVAR 46

RESULT 13
US-08-555-678-24
; Sequence 24, Application US/08555678
; Patent No. 5763174
; GENERAL INFORMATION:
; APPLICANT: Nishikura, Kazuko
; TITLE OF INVENTION: RNA Editing Enzyme and Methods
; TITLE OF INVENTION: of Use Thereof
; NUMBER OF SEQUENCES: 67
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Cntr, P.O. Box 457
; CITY: Spring House
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/555,678
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/197,794
; FILING DATE: 17-FEB-1994
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: WST49CUSA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-540-9206
; TELEFAX: 215-540-5818
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 71 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-457-459-24

Query Match 0.4%; Score 7; DB 1; Length 71;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 283 NGKEVAR 289
DB 40 NGKEVAR 46
```

Best Local Similarity 51.5%; Pred. No. 2.5e-65; Matches 286; Conservative 81; Mismatches 155; Indels 33; Gaps 9;

QY 5 GQEFLLANDIDNSNPVQAEQNLWYLLNFGTITANNDQANFDSVRVDA PDNIDADLMN 64
 Db 458 GQEFLLANDIDNSNPVQAEQNLWYLLNFGTITANNDQANFDSVRVDA PDNIDADLMN 64
 QY 65 IQADYFNAAGMD--SDAVSNKHINILEWHDADPEYFNKIGNPQMTDDTKNSLHNG-L 122
 Db 518 IASDYFKAKYGAQSQDAIKHLSLEAWSHNDAYNEDTKGAQLPMDPMLALVYSL 577
 QY 123 SDATNRWGLDAIVHOSLADRENSTENVIPYSPVRAHDNNSQDQIONAIRD-----VT 177
 Db 578 RPIGNRSVGEPLISNLSNDRSESGKNSKRMANYAFVRAHDSVQSIIGQIIKNEINPOST 637
 QY 178 GNDYHFTTFEDQKGDAYIODQNSVKKYNYLNIPASVAILLTNKTIPRYVYGDLYTD 237
 Db 638 G---NFTLDEMKAFFIYNKDMRSANKQYQYINIPSAVALMLTHKDTVPVYIGDMITD 694
 QY 238 GGOYMEHOTRYDITLNLKSRVYVAGGQSMQTSVG--GNNN-----ILTSVRYGKA 290
 Db 695 DGQYMAQKSPYDAIETLLKGRIRYVAGGQDMKVNYIGVNTNGHDAAGVLTSVRYGTA 754
 QY 291 MTATDGTDETRTQGIQVYVSNTPNKLGVNDKVVYVHGAHKNQYRAVLTITDGVIN 350
 Db 755 NSASDTGTAETRNQGMVIVSNQPALRLTSN--LTINMGAAHRNQAYRPLLTITDGVAT 812
 QY 351 YTSQDQAP--VAMTDENGDLVLSHNLVYNGKEADTA VQGVANPDVSGYLAVVVPVGAS 408
 Db 813 YLWDSANGIVKTDGNGMLTFSANE-----IRGINFPQDGYLAVVVPVGAS 860
 QY 409 DNQDARTAPSTERNKNSAYRTNAAFDSNVIFEAFNSFVYTPPKESERANVRIAQADPF 468
 Db 861 ENQDVRVAPSKENKSGLVYESNAALDSQVYEGFSNFQDFVQNPQYTNKKIAENANLF 920
 QY 469 ASIGFTSFEMAPQYNSKORTFLDSTIDNGYAFTRDYDGLGMSPNKYGYDEDLRNAIQAL 528
 Db 921 KSWGITSFEPAPQYVSSDDGSLDSVQNGYAFTRDYDGLGMSKDNKYGLADLKAALKSL 980
 QY 529 HKAGLOVMADWVDPQ 543
 Db 981 HAVGISAIADWVDPQ 995

RESULT 15
 Q55264 PRELIMINARY; PRT; 1449 AA.
 AC Q55264;
 DT 01-NOV-1996 (TRENBLrel. 01, Created)
 DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
 DE GLUCOSYLTRANSFERASE PRECURSOR.
 GN GPEL.
 OS Streptococcus salivarius.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1304;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95122197; PubMed=7822030;
 RA Simpson C.L., Giffard P.M., Jacques N.A.;
 RT "Streptococcus salivarius ATCC 25975 possesses at least two genes
 coding for primer-independent glucosyltransferases.";
 RL Infect. Immun. 63:609-621(1995).
 DR EMBL: L35495; AAC41412.1; -
 DR InterPro; IPR002479; CW_binding.
 DR InterPro; IPR003318; Glyco_hydro_70.
 DR Pfam; PF01473; CW_binding_1; 8.
 DR Pfam; PF02324; Glyco_hydro_70; 1.
 DR Signal; Transferase.
 FT SIGNAL 1 35 POTENTIAL.
 FT CHAIN 36 1449 GLUCOSYLTRANSFERASE.
 FT SEQUENCE 1449 AA; 159984 MW; DD62F07306E86A46 CRC64;

Query Match 49.4%; Score 1407.5; DB 2; Length 1449;
 Best Local Similarity 51.5%; Pred. No. 2.5e-65;
 Matches 286; Conservative 81; Mismatches 155; Indels 33; Gaps 9;

QY 5 GQEFLLANDIDNSNPVQAEQNLWYLLNFGTITANNDQANFDSVRVDA PDNIDADLMN 64
 Db 458 GQEFLLANDIDNSNPVQAEQNLWYLLNFGTITANNDQANFDSVRVDA PDNIDADLMN 64
 QY 65 IQADYFNAAGMD--SDAVSNKHINILEWHDADPEYFNKIGNPQMTDDTKNSLHNG-L 122
 Db 518 IASDYFKAKYGAQSQDAIKHLSLEAWSHNDAYNEDTKGAQLPMDPMLALVYSL 577
 QY 123 SDATNRWGLDAIVHOSLADRENSTENVIPYSPVRAHDNNSQDQIONAIRD-----VT 177
 Db 578 RPIGNRSVGEPLISNLSNDRSESGKNSKRMANYAFVRAHDSVQSIIGQIIKNEINPOST 637
 QY 178 GNDYHFTTFEDQKGDAYIODQNSVKKYNYLNIPASVAILLTNKTIPRYVYGDLYTD 237
 Db 638 G---NFTLDEMKAFFIYNKDMRSANKQYQYINIPSAVALMLTHKDTVPVYIGDMITD 694
 QY 238 GGOYMEHOTRYDITLNLKSRVYVAGGQSMQTSVG--GNNN-----ILTSVRYGKA 290
 Db 695 DGQYMAQKSPYDAIETLLKGRIRYVAGGQDMKVNYIGVNTNGHDAAGVLTSVRYGTA 754
 QY 291 MTATDGTDETRTQGIQVYVSNTPNKLGVNDKVVYVHGAHKNQYRAVLTITDGVIN 350
 Db 755 NSASDTGTAETRNQGMVIVSNQPALRLTSN--LTINMGAAHRNQAYRPLLTITDGVAT 812
 QY 351 YTSQDQAP--VAMTDENGDLVLSHNLVYNGKEADTA VQGVANPDVSGYLAVVVPVGAS 408
 Db 813 YLWDSANGIVKTDGNGMLTFSANE-----IRGINFPQDGYLAVVVPVGAS 860
 QY 409 DNQDARTAPSTERNKNSAYRTNAAFDSNVIFEAFNSFVYTPPKESERANVRIAQADPF 468
 Db 861 ENQDVRVAPSKENKSGLVYESNAALDSQVYEGFSNFQDFVQNPQYTNKKIAENANLF 920
 QY 469 ASIGFTSFEMAPQYNSKORTFLDSTIDNGYAFTRDYDGLGMSPNKYGYDEDLRNAIQAL 528
 Db 921 KSWGITSFEPAPQYVSSDDGSLDSVQNGYAFTRDYDGLGMSKDNKYGLADLKAALKSL 980
 QY 529 HKAGLOVMADWVDPQ 543
 Db 981 HAVGISAIADWVDPQ 995

RESULT 16
 Q48756 PRELIMINARY; PRT; 1290 AA.
 ID Q48756
 AC Q48756;
 DT 01-NOV-1996 (TRENBLrel. 01, Created)
 DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
 DE DEXTRANSUCRASE.
 OS Leuconostoc mesenteroides.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
 OC Leuconostoc.
 OX NCBI_TaxID=1245;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN=NRRL B1299;
 RX MEDLINE=97136686; PubMed=8982063;
 RA Monchois V., Willemot R.M., Renaud-Simeon M., Croux C., Monsan P.;
 RT "Cloning and sequencing of a gene coding for a novel dextranucrase
 from Leuconostoc mesenteroides NRRL B-1299 synthesizing only alpha (1-
 6) and alpha (1-3) linkages.";
 RL Gene 182:23-32(1996).
 DR EMBL: U39181; BAB40875.1; -
 DR InterPro; IPR002479; CW_binding.
 DR InterPro; IPR003318; Glyco_hydro_70.
 DR Pfam; PF01473; CW_binding_1; 11.
 DR Pfam; PF02324; Glyco_hydro_70; 1.


```
SQ SEQUENCE 1290 AA; 145590 MW; 3555C2E96B749FAA CRC64;

Query Match
Best Local Similarity 48.5%; Score 1383; DB 2; Length 1290;
Matches 279; Conservative 84; Mismatches 161; Indels 24; Gaps 7;

QY 5 GQEFLLANDIDNSNPVQAEQLNMLYLLNFETITANNQANFDSVRVDAPDNIDADLMN 64
DB 238 GRELLANDVDNSNPVQAEQLNMLYLLNFETITANNQANFDSVRVDAPDNIDADLIQ 297
QY 65 IQADYFNAAYGM-DSDAYSNKHINLEDDNHADPEYFNKIGNPOLTMDDTTKNSLHG 121
DB 298 IAGEYAKAAYGVDK--MTRERINIVQFWKTGEMKIQTMKHNATSKLSMDFPLHLAIKTA 355
QY 122 LSDATN-RGLDAIVHQSLADRENNSTENVIPNYSFVRAHDNNSQDIQNAIRD----- 175
DB 356 LAMPNCKRSGLEPTEHSLVKRTDDKENVAOPNYSFIRAHSEVQTIADIIKOKINPA 415
QY 176 VTGKDYHTFFDEQKIDAYIQDQNSTVKKYNLYNPASAYAILLTKNKTIPRYVYGDLY 235
DB 416 STGLD-STVTLQIKQAFDIYNADELKADKVTPYPNIPASAYAILLTKNKTIPRYVYGMF 474
QY 236 TDGQYMEHGTFRYDFTLNLKSRVYVAGGQSMQTMVSGNNILTSVRYGKGAMTAD 295
DB 475 TDGQYNAQSPYQAIQDALLKARIKAYAGGQTMKNYPDQSGVMTSVRYGKGAMTASD 534
QY 296 TGTDEFTQIGVWVSVNTPNLKLGVNDKVLVHMGAAHKNQOYRAAVLTITDGVINVT 355
DB 535 SGQETRYQIGLGVNRRPDLKLSKDEKVMGMDGAHKNQDYRPVLLTTKSLKVVSTDA 594
QY 356 GAPVMTDENGDIYLSHNLVYNGKEADTAVOGYANPDVSGYLAWVVPVGSADQART 415
DB 595 NADPVRTDANGQLTF-----KAD-MVYGVNDPQVSGYIAAWVVPVGSADQART 642
QY 416 APTEKNSGNSAVRTNAADSNVFEAFSNFTVPTPKESERANVRVIAQNAADPFA 475
DB 643 KSETQSTGDSVYHSNALDSQVYIEGFSNFQDPTTPEFTNIIKIAQNVNLEKDNGLIS 702
QY 476 FEMAPVNSKDRFTLDTSTIDNGYAFDTRYDLGMSBNPKYGTDEDLNAIQALHAGLOV 535
DB 703 FEMAPVYRSDSKSFLQAYQYNGYAFDTRYDYGYNTKYGTADNLIDALRALHGGQIOA 762
QY 536 MADWVPDQ 543
DB 763 INDWVPDQ 770

RESULT 17
Q9LCH3 PRELIMINARY; PRT; 1575 AA.
AC Q9LCH3;
DT 01-OCT-2000 (TREMELrel. 15, Created)
DT 01-OCT-2000 (TREMELrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMELrel. 17, Last annotation update)
DE GLUCOSYLTRANSFERASE.
GN GTRF.
OS Streptococcus oralis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1303;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC10557;
RX MEDLINE=20231779; PubMed=10768934;
RA Fujiwara T., Hoshino T., Ooshima T., Sobue S., Hamada S.;
RA "Purification, characterization, and molecular analysis of the gene
RT encoding glucosyltransferase from Streptococcus oralis.";
RL Infect. Immun. 68:2475-2483(2000).
DR ENBL; AB025228; BAA95201.1;
DR InterPro; IPR002479; CW-binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 17.
```

```
DR Pfam; PF02324; Glyco_hydro_70; 1.
RW Transferrase.
SQ SEQUENCE 1575 AA; 176792 MW; 772A26E4D7C2B543 CRC64;

Query Match
Best Local Similarity 47.4%; Score 1351.5; DB 2; Length 1575;
Matches 272; Conservative 91; Mismatches 164; Indels 27; Gaps 6;

QY 5 GQEFLLANDIDNSNPVQAEQLNMLYLLNFETITANNQANFDSVRVDAPDNIDADLMN 64
DB 468 GYEFLANDFDSNPVQAEQLNMLYLLNFETITANNQANFDSVRVDAPDNIDADLIQ 527
QY 65 IQADYFNAAYGM-DSDAYSNKHINLEDDNHADPEYFNKIGNPOLTMDDTTKNSLHG-L 122
DB 528 TASDYFRKYKVGSEEEAIIKLSLEAWSNDPDYNTKGAQLAIDNKLRLSLYSFM 587
QY 123 SDATNWKGLDAIVHQSLADRENNSTENVIPNYSFVRAHDNNSQDIQNAIRD--VTGK 180
DB 588 RMLSTRSGVEPTITSLANDRSSEKKNGERMANYIFRAHDSVQIVADIIRENINPWT 647
QY 181 YHTFFDEQKIDAYIQDQNSTVKKYNLYNPASAYAILLTKNKTIPRYVYGDLYTDGQ 240
DB 648 GLTFTMDLQKAFKTYNEDMRKADKKTQFNIPTAHALMLSNKDSITRVYGDLYTDGQ 707
QY 241 YMEHGTTRYDFTLNLKSRVYVAGGQSMQTMVSG-----GNNILTSVRYGKGAMT 292
DB 708 YMEKSPYHDAIDALLKARIKAYAGGQDMKVYMGVPRADKWSYNGILTSVRYGTGANE 767
QY 293 ADTGTDEFTQIGVWVSVNTPNLKLGVNDKVLVHMGAAHKNQOYRAAVLTITDGVINVT 352
DB 768 ATDEGTAETRTQGMVIAVSNPNLXNLNEDKLVQVNGAAHKNQYRPVLLTTKDGISRYL 827
QY 353 SDQGAIVAM---TDENGDIYLSHNLVYNGKEADTAVOGYANPDVSGYLAWVVPVGSAD 409
DB 828 TDEEVFQSLKKTDTANGILTFDND-----IAGYSNVQVSGYLAWVVPVGA 875
QY 410 NODARTAPSTENSGNSAVRTNAADSNVFEAFSNFTVPTPKESERANVRVIAQNAADPFA 469
DB 876 DQDARTATSKKNASQVYESSAALDSQVYIEGFSNFQDPTTQNKVIAKNVNLFK 935
QY 470 SLGFTSFEMAPVNSKDRFTLDTSTIDNGYAFDTRYDLGMSBNPKYGTDEDLNAIQALH 529
DB 936 EMGVTSFELPPQVYSSQDGTFLDSIIQNGYAFEDRYDMAMSKNNKYGSLKDLNLRALH 995
QY 530 KAGLQVMAWVVPDQ 543
DB 996 SVNIIQAIADWVDPQ 1009

RESULT 18
Q54178 PRELIMINARY; PRT; 1577 AA.
AC Q54178; Q54247;
DT 01-NOV-1996 (TREMELrel. 01, Created)
DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMELrel. 17, Last annotation update)
DE GLUCOSYLTRANSFERASE.
GN GTRF.
OS Streptococcus gordonii challis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=29350;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=CHALLIS;
RX MEDLINE=96157084; PubMed=8586195;
RA Vickerman M.M., Sulavik M.C., Clewell D.B.;
RA "Molecular analysis of Streptococcus gordonii glucosyltransferase
RT phase variants.";
RL Dev. Biol. Stand. 85:309-314(1995).
DR [2]
RP SEQUENCE OF 1-96 FROM N.A.
```


GN	GTFs.	DE	GLUCOSYLTRANSFERASE I (EC 2.4.1.5) (GTF) (DEXTRANSUCRASE).	QY	5	GOEFLLANDIDNSNPVQAEOLNLYLLNPGTITANNQANFDSVRVDAPONDADLN 64
OS	Streptococcus criceti.	DE	(SUCROSE 6-GLUCOSYLTRANSFERASE).	Db	423	GFDFELLANDVLSNPVQAEOLNLYLLNPGTITANNQANFDSVRVDAPONDADLN 482
OG	Plasmid pAM1.	GN	GTFJ.	QY	65	ITADYFNAAYGMD-SDAVSNKHINLEDDNHADPEYFNKIGNPQLTMDDTIKNSLHGLS 123
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;	OS	Streptococcus salivarius.	Db	483	LYTNRYFYYGVNKSSEANALAHLSVLEANSLENDNHYNDKTDGAALAMENKQRLALLFLS 542
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;	OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;	QY	124	-----DATNRWGLDAIVHOSLADRENNSTENVV 151
OC	Streptococcus	OC	Streptococcus	Db	543	KPIKERTPAVSPLYNNFTNTQDEKTDNWKDQSKAINEDEGVVKOSTGKYNEKGD-A 601
OX	NCBI_TaxID=1333;	OC	Streptococcus	QY	152	IPAKSVFRAHNNSSQOIQNAI-RDVTGK-DYHTFTPEDEKQIDAYIQDNSTVKYKL 209
RN	[1]	OC	Streptococcus	Db	602	SGNVVFIHNDNNVQDIIAEI LKKEINPKSDGFTITDAEMKQAFELYNKDMLSSDKKYL 661
RP	SEQUENCE FROM N.A.	OC	Streptococcus	QY	210	YNTPASVAILLTKNDTIPRYVYGDLYTDGQYMEHQRYVYDTLTNLLKSRVKYVAGQSM 269
RC	STRAIN=HS-6;	OC	Streptococcus	Db	662	NNTPAAVAVMLQNMETITRYVYGDLYTDGQYMEHQRYVYDTLTNLLKSRVKYVAGQSM 721
RA	Inoue M., Fukui K., Miyagi A.;	OC	Streptococcus	QY	270	QT-----MSVGGNNILTSRYGKGMATATDT-GTDETRTGIGVYVNTPN 315
RT	Submitted (MAR-1999) to the EMBL/Genbank/DBJ databases.	OC	Streptococcus	Db	722	RSYWLPTDGMKNDNSDELRYTNEVYTSVRIGKDKLMDANDTEGSKYTSQVTLVANNPK 781
DR	EMBL; AB026123; SAA77236.1; .	OC	Streptococcus	QY	316	LKLGVNDKVVLMHGAHKNQYRAAVLTFTDGVNNTSDQAPVA-----MTDENGDIYLS 371
DR	InterPro; IPR002479; CW_binding_70.	OC	Streptococcus	Db	782	LNLQSAKLVEMCKIHANQYRALIVGTADGINKFTSDADATAAGYVVKETDSNGVLTFG 841
DR	InterPro; IPR003318; Glyco_hydro_70.	OC	Streptococcus	QY	372	SHNLVNGKEADTAVOGYANPQVSGYLAVVVPVVGASDNQDARTAPSTE-KNSGNSAIRT 430
DR	Pfam; PF01473; CW_binding_1; 10.	OC	Streptococcus			
DR	Pfam; PF02324; Glyco_hydro_70; 1.	OC	Streptococcus			
KW	Plasmid.	OC	Streptococcus			
SQ	SEQUENCE 1338 AA; 148558 MW; 0A90C8E10E15D99B CRC64;	OC	Streptococcus			
Query Match						43.8%; Score 1249.5; DB 2; Length 1338;
Best Local Similarity						47.5%; Pred. No. 3.9e-57;
Matches 262; Conservative 84; Mismatches 179; Indels 27; Gaps 8;						
QY	3	YRGOEFLLANDIDNSNPVQAEOLNLYLLNPGTITANNQANFDSVRVDAPONDADLN 62				
Db	357	YAGYELLANDVDSNPVQAEOLNLYLLNPGTITANNQANFDSVRVDAPONDADLN 416				
QY	63	MNTAODYFNAAYGMD-SDAVSNKHINLEDDNHADPEYFNKIGNPQLTMDDTIKNSLHGLS 121				
Db	417	LQIQDYKAKYGVQDNKNAIDHLSLEAMSGNDNDYKQDNFSLDSQSYMLAA 476				
QY	122	LSDATNRWGLDAI---VHQSADRENNSTENVVIPYKSVFRAHNNSSQOIQNAIRDVTG 178				
Db	477	FAYPASQGRGDIYISLLPKVGLKDRYAKNGNPV-PNVVFIHNDNNVQDIIAEI LKKEINPKSDGFTITDAEMKQAFELYNKDMLSSDKKYL 535				
QY	179	K----DYHTFTPEDEKQIDAYIQDNSTVKKYLNIPASVAILLTKNDTIPRYVYGD 234				
Db	536	KTNADGTLNITLDDLNKAFDIYNDMDKAVQKYPNPLPWAYAMWLNQKDTVRYVYGD 595				
QY	235	YTDGQYMEHQTRYDITLNLKSRVYVAGGSMQTMNSVGGNNILTSRYGKGMATAT 294				
Db	596	YTDGQYMEHQTRYDITLNLKSRVYVAGGSMQTMNSVGGNNILTSRYGKGMATAT 654				
QY	295	DIGTDETRTGQIGVVSNTPNLKLGVNDKVVLMHGAHKNQYRAAVLTFTDGVNNTSD 354				
Db	655	DAGTAETRTSGMAVLINNKPNFQ---SYNGLTLMGAAHKNQYRAVPELSTKDSGIATYLD 712				
QY	355	QGA---PVMTDENGDIYLSHNLVNGKEADTAVOGYANPQVSGYLAVVVPVVGASDNQ 411				
Db	713	SDVSSNQYKTDGGRNLFSAEL-----RSVANQVQSMQIYVVPVVGASDNQ 760				
QY	412	DARTAPSTPEKNSGNSAYRTNAAFDSNVIFEAFSNVYTPTKESERANVRVIAONAF 471				
Db	761	DVRVAPNTNRNNSNIYTSQDALSQVYIGFSNFQAFATPEQYTNVAVIAKNADLFKSW 820				
QY	472	GFTSEMAPQVNSKDRFTDSTDNGVAFTRDYDLGSEPNKYGTDEDLNNAQLHKA 531				
Db	821	GITQFEMAPQVSSDEDTFLDSVVLNGYAFSDRYDLAMSKNNKYGSKEDLANAKGLHNA 880				
QY	532	GLOVHADAVPDQ 543				
Db	881	GIKVLSMDVPDQ 892				
RESULT	21					
Q00600						
ID	Q00600	PRELIMINARY;	PRT;	1518	AA.	
AC	Q00600;					
DT	01-NOV-1996 (Tremblrel. 01, Created)					
DT	01-NOV-1996 (Tremblrel. 01, Last sequence update)					
DT	01-JUN-2001 (Tremblrel. 17, Last annotation update)					


```
Db 411 GYEFLSLNDIDNSPVVQAEMLNLYTFMNGQVIFGDKDKDAHFDDGIRYDAVDNVSVM 470
QY 63 MNIAQDYFNAAYGM-DSDAVSNKHINILEDNWHAADPEVFNKIGNPQMTMDT IKNSLNHG 121
Db 471 LQLVSSYKAAKYKNESEARLAWISLEASHNDPPYVNEHTAALSDNGSLRSIVHG 530
QY 122 LSDATNRWGLDAIVHQSAD-----RENNSTENVVPIPNYSFVRADNNSQOI 169
Db 531 LTRPVNKGTA-RNASMKDLINGYFGLSNRAEVTSTYDQLCFATYLFVRADS-----EV 585
QY 170 QNAIDVYTK-----DYHTTFDEQKIDAYIQDQNSTVKYKYNLYNIPASYAILTNK 223
Db 586 QTVIADII SKKIDPTDGTFTLQKQAFDIYNADMLKVDKEYTHSNIPAYAILMQTM 645
QY 224 DIPRVYVYGLYTDGQGYMEHQTRYDFTLNLSKRVYVAGQSMQMSVGG----- 276
Db 646 GAATRVYVGLYTDGQGYMAKSPYFDITILLKARPKYVAGGQTSYIHNLAGDGVSSAK 705
QY 277 -NNNTLTSVRKGGAMTADT-GTDETFRTQGVVVSNTPNKLGVDNKKVYLHMGAAHN 334
Db 706 DNKEVLVSRYGQDLMSKTDTEGGYKRGNSGMLTLIANNPDLKLADGETITVNMGAHN 765
QY 335 QYRAAVLTITDGVNYSQDAP-VAMTDENGDLVLSHNLVYNGKEADTAQGYANP 393
Db 766 QAYRELLLGTEKGISSLSDSTKIKYKTDAG-----NLVFTADE-----KGFKTIV 813
QY 394 DVSGLVAVVYPVYASDNQDARTAPSTEK-NSGNSAYRTNAAAFDSNVIFEAFSNFYPTK 452
Db 814 DMSGYLWVYVYVAGDDQNVLAKEPTKAYKSGDKYVSSAALAEQVYVIEGFSNFQDFYKE 873
QY 453 ESERANVRIAQNAADFFASLGFTSFEMAFQYNSKDRFLDSTIDNGYAFTRDYDLGMEP 512
Db 874 DSOYTNKLIANAADLFKSGITSFBIATQYVSSKGTDLDSIENGYAFTRDYDFAMSKN 933
QY 513 NKYGTEDELNAIQALHKAQVADWVDPQ 543
Db 934 NKYGSKEDLRDALKAHKQGIQVADWVDPQ 964

RESULT 24
Q93148 ID Q93148 PRELIMINARY: PRT: 501 AA.
AC Q93148;
DT 01-DEC-2001 (TremBLrel. 19, Created)
DT 01-DEC-2001 (TremBLrel. 19, Last sequence update)
DE AMYLASE.
GN AMYK38.
OS Bacillus sp. KSM-K38.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=129736;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-KSM-K38;
RA Hayashi Y.;
RT "Isolation of a new Bacillus alpha-amylase.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB051102; BAB71820.1; -.
SQ SEQUENCE 501 AA; 57485 MW; 1240F46739A5CC11 CRC64;
```

Query Match 5.9%; Score 169; DB 2; Length 501;
Best Local Similarity 21.6%; Pred. No. 0.22;
Matches 75; Conservative 44; Mismatches 106; Indels 122; Gaps 15;

```
QY 7 EFLANDIDNSNPVQAEQLNWLNYLLNFTITANNQDANFDSVRVADPNIDADLWNTA 66
Db 215 DYLLGNSIDFSHPVQDELKGSWF-----TDELQDGRILDA----- 253
QY 67 QDYFNAAYGMSDAVSNKHINI-LTDW-----NHADPEVF-----NKIGNPQMTMD 112
```

```
Db 254 -----IKHIPFWVTSWVRQRNEADQDLFVVGVEYWKDDVGAEFLVDE 297
QY 113 TIKNSLNHGLSDATNRWGLDAIVHQSADRENNSSTENVIP-NYSFVRADNNSQOIQON 171
Db 298 M-----EMSLFD-----VPLNINFRASQGGSGYDMRN 327
QY 172 AIRD--VTGKDYHTFFEDQ-----KGIDAYIQDQNSTVKYKYNLYNIPASYAILTNK 284
Db 328 ILKGLVNEAHPMAHTVFDVNDHDPGESLESWAD-----WFKPLAYATILYRREG 377
QY 225 TTPRVYVYGLYTDGQGYMEHQTRYDFTLNLSKRVYVAG-----GQSMQTM 272
Db 378 GYPNVYVDYI--GIPNDNISAKMDIDELLDARONYAYGTQHDYFDHMDVYVGTREGS 434
QY 273 SVGGNNILSVRYGKA-----WTATDTGTDETRTQGIQGVVVS 311
Db 435 SSRENSGLATIMSGPGGSKWYVGRQNAQTWTDLTGNNGASVTIN 481

RESULT 25
O68875 ID O68875 PRELIMINARY: PRT: 486 AA.
AC O68875;
DT 01-AUG-1998 (TremBLrel. 07, Created)
DT 01-AUG-1998 (TremBLrel. 07, Last sequence update)
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE INTRACELLULAR A-AMYLASE.
GN AMY.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-L711;
RA MEDLINE-98389691; PubMed-9721315;
RA Simpson C.L., Russell R.R.B.;
RT "Intracellular alpha-amylase of Streptococcus mutans.";
RL J. Bacteriol. 180:4711-4717(1998).
DR EMBL; AF055987; AAC35010.1; -.
DR HSSP; P06278; IVJS.
DR InterPro; IPR000461; Alpha_amylase.
DR Pfam; PF00128; alpha-amylase; 1.
SQ SEQUENCE 486 AA; 56351 MW; 18DB93072F428938 CRC64;
```

Query Match 5.9%; Score 168; DB 2; Length 486;
Best Local Similarity 22.5%; Pred. No. 0.24;
Matches 80; Conservative 63; Mismatches 121; Indels 92; Gaps 15;

```
QY 7 EFLANDIDNSNPVQAEQLNWLNYLLNFTITANNQDANFDSVRVADPNID-----ADL 62
Db 197 DYLMYDDIDFRHPEVQEHLRDWVAMFL-----ETSGVGGFRLDAIKHIDKTFMAQF 247
QY 63 MNIAQDYFNAAYGMSDAVSNKHINILEDNWHAADPEVFNKIGNPQMTMDT IKNSLNHG 122
Db 248 IRYIREHLKADLYVFGYWKDSDHEDI-TDYLHSDVLQFDLI----- 287
QY 123 SDATNRWGLDAIVHQSADRENNSSTENVVPIPNYSFVRADNNSQOIQNAIRDVTKD 180
Db 288 -----DVMLHMSLFEAGQKGSDFDLSTILDLSLAKSHPDFA-----VTFVD 328
QY 181 YHTTFDEQKIDAYIQDQNSTVKYKYNLYNIPASYAILTNKDTIPRVYVYGLYTDGQ 240
Db 329 NH-----DSQG-----QALESTVAEFK---PLAYGLILLQEGIPCVFDYIGISGE 375
QY 241 YMEHOTRYDITLNLSKRVYVAGQSMQMSVGGNNNLT SVRYGKAMTATDGTDE 300
Db 376 FA--QESFQTVLDKLLY IROYHYVYGSKKI-----ILNM---PNCIGWTCLG-DE 418
QY 301 TRTQIGVYVVSNTPNKLGVDNKKVYLHMGAAHNQOYRAAVLTITDGVNYSQOG 356
Db 419 EHPDGVAVIISN-----GEANCKRMNMGFEFRNKFVVDYLNKTEEVI--LDDQG 466
```

Search completed: August 12, 2002, 08:03:55
Job time: 247 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

4 protein - protein search, using sw model

in on: August 12, 2002, 08:00:38 ; Search time 29.85 Seconds
(without alignments)
993.524 Million cell updates/sec

title: US-09-995-749a-2_COPY_1515_1781
perfect score: 1430
sequence: 1 IYNLPKGEVATVTRVDRGN.....SDAEYPTSTDVGKMLDQNK 267

scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

searched: 747574 seqs, 111073796 residues

total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Maximum Match 0%
Listing first 100 summaries

Database : A_Geneseq_032802.*
1: /SIDS1/gcgdata/hold-geneq/geneq-emb1/AA1980.DAT.*
2: /SIDS1/gcgdata/hold-geneq/geneq-emb1/AA1981.DAT.*
3: /SIDS1/gcgdata/hold-geneq/geneq-emb1/AA1982.DAT.*
4: /SIDS1/gcgdata/hold-geneq/geneq-emb1/AA1983.DAT.*
5: /SIDS1/gcgdata/hold-geneq/geneq-emb1/AA1984.DAT.*
6: /SIDS1/gcgdata/hold-geneq/geneq-emb1/AA1985.DAT.*
7: /SIDS1/gcgdata/hold-geneq/geneq-emb1/AA1986.DAT.*
8: /SIDS1/gcgdata/hold-geneq/geneq-emb1/AA1987.DAT.*
9: /SIDS1/gcgdata/hold-geneq/geneq-emb1/AA1988.DAT.*
10: /SIDS1/gcgdata/hold-geneq/geneq-emb1/AA1989.DAT.*
11: /SIDS1/gcgdata/hold-geneq/geneq-emb1/AA1990.DAT.*
12: /SIDS1/gcgdata/hold-geneq/geneq-emb1/AA1991.DAT.*
13: /SIDS1/gcgdata/hold-geneq/geneq-emb1/AA1992.DAT.*
14: /SIDS1/gcgdata/hold-geneq/geneq-emb1/AA1993.DAT.*
15: /SIDS1/gcgdata/hold-geneq/geneq-emb1/AA1994.DAT.*
16: /SIDS1/gcgdata/hold-geneq/geneq-emb1/AA1995.DAT.*
17: /SIDS1/gcgdata/hold-geneq/geneq-emb1/AA1996.DAT.*
18: /SIDS1/gcgdata/hold-geneq/geneq-emb1/AA1997.DAT.*
19: /SIDS1/gcgdata/hold-geneq/geneq-emb1/AA1998.DAT.*
20: /SIDS1/gcgdata/hold-geneq/geneq-emb1/AA1999.DAT.*
21: /SIDS1/gcgdata/hold-geneq/geneq-emb1/AA2000.DAT.*
22: /SIDS1/gcgdata/hold-geneq/geneq-emb1/AA2001.DAT.*

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	582.5	40.7	2057	21	AA10667 L. mesenteroides a
2	439.5	30.7	1592	14	AA32295 Glucosyltransferase
3	418	29.2	1577	17	AA391047 Alpha-D-glucosyltr
4	134.5	9.4	591	19	AAW50599 Streptococcus pneu
5	134	9.4	659	21	AAV81517 Streptococcus pneu
6	132	9.2	319	21	AAV81660 Streptococcus pneu
7	129	9.0	678	21	AAV81667 Streptococcus pneu
8	128	9.0	569	19	AAW61214 Streptococcus pneu
9	126	8.8	214	22	AAV98217 Peptostreptococcus
10	125	8.7	628	21	AAV81659 Streptococcus pneu
11	125	8.7	2710	17	AAV95016 C. difficile toxin

12	125	8.7	2710	19	AAW68387
13	120.5	8.4	453	20	AAV06593
14	120.5	8.4	606	20	AAV43219
15	119.5	8.4	239	20	AAV25380
16	119.5	8.4	239	20	AAV02636
17	119.5	8.4	292	20	AAV25379
18	119.5	8.4	292	20	AAV02635
19	119.5	8.4	390	20	AAV25381
20	119.5	8.4	390	20	AAV02637
21	119	8.3	445	20	AAV06592
22	117.5	8.2	929	18	AAW14593
23	117	8.2	862	20	AAW87971
24	117	8.2	866	21	AAV30543
25	117	8.2	866	21	AAV30546
26	116.5	8.1	296	21	AAV16697
27	116	8.1	631	21	AAV30544
28	115.5	8.1	511	17	AAV95012
29	115.5	8.1	511	17	AAV75371
30	115.5	8.1	608	17	AAV95013
31	115.5	8.1	609	17	AAV95018
32	115.5	8.1	611	17	AAV75372
33	115.5	8.1	2366	17	AAV95011
34	115.5	8.1	2366	19	AAW68388
35	114.5	8.0	539	20	AAV43393
36	114.5	8.0	631	18	AAW46444
37	114.5	8.0	663	20	AAV49232
38	114.5	8.0	663	20	AAV32104
39	114.5	8.0	663	20	AAV43395
40	113.5	7.9	929	20	AAV43384
41	113	7.9	811	17	AAV95017
42	113	7.9	812	17	AAV95014
43	112	7.8	261	19	AAW61213
44	112	7.8	333	21	AAV81658
45	110.5	7.7	2314	22	AAV80136
46	109.5	7.7	320	20	AAW81969
47	109.5	7.7	457	20	AAW87970
48	109.5	7.7	1315	20	AAV08642
49	108	7.6	471	19	AAW61216
50	108	7.5	490	21	AAV81745
51	107.5	7.5	626	21	AAV95782
52	106.5	7.4	316	19	AAV61215
53	106.5	7.4	341	21	AAV81656
54	104.5	7.3	405	19	AAV85939
55	104.5	7.3	1349	22	AAU34402
56	104.5	7.3	1349	22	AAU37542
57	103	7.2	179	21	AAV71283
58	103	7.2	453	19	AAW5081
59	103	7.2	655	20	AAV49226
60	103	7.2	655	20	AAV32099
61	103	7.2	694	21	AAV81653
62	102	7.1	448	21	AAV71285
63	101.5	7.1	1359	12	AAV10562
64	101.5	7.1	1362	12	AAV10560
65	101.5	7.1	1362	12	AAV10561
66	101.5	7.1	1362	12	AAV10557
67	101.5	7.1	1362	12	AAV10558
68	101.5	7.1	1362	12	AAV10559
69	101.5	7.1	1362	12	AAV10563
70	101.5	7.1	1368	12	AAV10541
71	101.5	7.1	1374	12	AAV10940
72	100.5	7.0	210	19	AAW5080
73	98.5	6.9	2053	22	AAV69135
74	98	6.9	368	19	AAV55100
75	98	6.9	1584	20	AAV33727
76	97.5	6.8	1262	18	AAV13505
77	97.5	6.8	1644	18	AAV13504
78	97	6.8	182	21	AAV81657
79	96.5	6.7	416	22	AAV59569
80	96	6.7	745	22	AAV81652
81	96	6.7	867	22	AAV98887
82	95.5	6.7	437	16	AAV75507
83	95.5	6.7	641	19	AAW61217
84	95.5	6.7	1095	21	AAV801835

Clostridium diffi
CLYTA-WAGE-3-His f
E. rhusiopathiae s
HPV fusion protein
CLYTA-E7-His prote
HPV fusion protein
CLYTA-B6-His prote
HPV fusion protein
CLYTA-B6E7-His pro
CLYTA-WAGE-1-His f
Streptococcus pneu
Toxin A immunogeni
A recombinant prot
A recombinant prot
Bacteriophage Dp-1
A recombinant prot
C. difficile toxin
C. difficile toxin
C. difficile toxin
C. difficile toxin
C. difficile toxin
C. difficile toxin
C. difficile toxin
S. pneumoniae SpA
S. pneumoniae SpA
ChpA, a choline bi
ChpA of serotype R
Choline binding pr
S. pneumoniae chol
S. pneumoniae FspC
C. difficile toxin
C. difficile toxin
Streptococcus pneu
Streptococcus pneu
Streptococcus pneu
M. catarrhalis les
Toxin A immunogeni
Toxin A immunogeni
S. aureus SDRD pro
Streptococcus pneu
Streptococcus pneu
Erysipelothrix rhu
Streptococcus pneu
Streptococcus pneu
S. pneumoniae deri
Staphylococcus aur
Staphylococcus aur
Streptococcus pneu
Streptococcus pneu
Streptococcus pneu
CpA of serotype 4
Choline binding pr
Streptococcus pneu
Streptococcus pneu
Mutant protease (d
Mutant protease (K
Mutant protease (N
Mutant protease (A
Mutant protease (A
Mutant protease (A
Mutant protease (K
Mutant protease (d
Mutant protease (d
Streptococcus pneu
M. catarrhalis str
Streptococcus pneu
Streptococcus pneu
Streptococcus pneu
Streptococcus pneu
Drosophila melanog
Streptococcus pneu
E. coli growth and
Nicotiana glumabagi
Streptococcus pneu
Haemophilus infliue

85 95.5 6.7 1101 21 AAB01834 Haemophilus influe
86 94.5 6.6 382 21 AAB36303 BOT/A prototoxin
87 94.5 6.6 415 22 AAB04083 Botulism toxin C f
88 94.5 6.6 432 21 AAY77142 Native botulinum n
89 94.5 6.6 434 22 AAB04089 Botulism toxin hea
90 94.5 6.6 435 22 AAB04090 Botulism toxin hea
91 94.5 6.6 437 22 AAB04088 Botulism toxin hea
92 94.5 6.6 438 17 AAR95008 Type A neurotoxin
93 94.5 6.6 438 19 AAW68389 Clostridium botuli
94 94.5 6.6 438 21 AAY77134 Synthetic botulinu
95 94.5 6.6 445 19 AAW68391 Clostridium botuli
96 94.5 6.6 462 17 AAR95009 Type A neurotoxin
97 94.5 6.6 462 19 AAW68390 Clostridium botuli
98 94.5 6.6 619 15 AAR63437 Pneumococcal surfa
99 94.5 6.6 619 17 AAR87598 Pneumococcal surfa
100 94.5 6.6 619 17 AAR86911 Pneumococcal surfa

ALIGNMENTS

RESULT 1
D AAB10667 standard; Protein; 2057 AA.
X X AAB10667;
T 19-JAN-2001 (first entry)
E L. mesenteroides alternan sucrose protein.
W Alternan sucrose; glucosyltransferase; fructose; cosmetic; foodstuff;
W syrup.
X Leuconostoc mesenteroides.
S DE19905069-A1.
N 10-AUG-2000.
D 08-FEB-1999; 99DE-1005069.
F 08-FEB-1999; 99DE-1005069.
R (PLAN-) PLANTTEC BIOTECHNOLOGIE GMBH.
X (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
A Kossmann J, Welsh T, Quanz M, Knuth K;
I WPI: 2000-550294/51.
R N-PSDB; AAA97904.
X New nucleic acid encoding recombinant Leuconostoc mesenteroides
T alternan sucrose protein and methods of alternan and fructose
T production -
T Claim 1a; Page 30-36; 64pp; German.

X This invention describes a novel nucleic acid molecule (I) encoding an
X alternan sucrose (E.C. 2.4.1.140 - an enzyme, that belongs to the
C glucosyltransferase group) the recombinant, purified alternan sucrose
C gene is useful for the fermentative production of alternan (a
C carbohydrate) and/or fructose by secreting the enzyme into a
C saccharose-containing culture medium. Alternatively, the enzyme is
C contacted with a saccharose-containing solution. The alternan and/or
C fructose is then isolated from the medium. Cosmetic products or
C foodstuffs containing alternan can be produced. Recombinant production of
C alternan sucrose is advantageous as it provides a cost effective means of
C producing fructose for high fructose containing syrups, production of
C which previously has been achieved by costly production from maize
C starch. This sequence represents the Leuconostoc mesenteroides alternan
C sucrose protein which is described in the method of the invention.

SQ Sequence 2057 AA;
Query Match 40.7%; Score 582.5; DB 21; Length 2057;
Best Local Similarity 48.5%; Pred. No. 2.1e-44;
Matches 127; Conservative 34; Mismatches 86; Indels 15; Gaps 8;
QY 1 IYNLPGEVAVTVRVDGRGNWKAIIINNLYVNTTGGG-EYOKKYGAFIDKQKLYP 59
DB 1175 IYALPGKEVAVTVRVDGRGNWKAIIINNLYVNTTGGG-EYOKKYGAFIDKQKLYP 1234
QY 60 EFTTKKQYSTGVAIDPSOKITEWSAKYFNGTNIHRSYVYLKA-DGGQYVNLGTTTKQF 118
DB 1235 SLFKQNGYSTGPIQDAKTKLQWSAKYFNGTNIHRSYVYLKA-DGGQYVNLGTTTKQF 1294
QY 119 LPIQITGKKGQNEGFVAKGNDGNYFYFDLAGNMVAKNTFIEDSVGNWYFFDQDGRKVENK- 177
DB 1295 IPlql-qkdaqtgfisdaagvky-ysisgyakdtfiedgngnwyfyfdkgymvrsq 1351
QY 178 -----HFVDVDSYGEKGYFFFLKNGVSPRAGLVQTDN--GTYFFDNYGKMYRNQPINAGA 230
DB 1352 genpirtvetsvntngnyfmpngveIrkGf-gtdnsgnvyyfdagqgmvrdrkyinda 1410
QY 231 -MIYTLDENGKLKASVNSDAE 251
DB 1411 mnfyhlnvdtmsrglfkfdsd 1432
RESULT 2
AAR32925
ID AAR32925 standard; Protein; 1592 AA.
XX AAR32925;
AC AAR32925;
XX 28-JUN-1993 (first entry)
DT Glucosyltransferase I.
DE GT-1; Streptococcus; dental; caries.
XX Streptococcus sobrinus.
OS JP05023188-A.
PN 02-FEB-1993.
PD 25-JUL-1991; 91JP-0186592.
PF 25-JUL-1991; 91JP-0186592.
PR (FUKU/) FUKUI I.
XX (KATO/) KATO K.
PA WPI: 1993-079449/10.
XX DR N-PSDB; AAQ37760.
DR DNA sequence glucosyl:transferase-1 - comprises Streptococcus
XX sobrinus DNA sequence with at least one nucleotide added or
PT deleted
PT Claim 13; Page 15; 29pp; Japanese.
XX The DNA sequence from Streptococcus sobrinus strain 6715 encodes
CC glucosyltransferase-I (and mutants). The DNA was obtd. by treating
CC S. sobrinus 6715 with mutanolysin, extracting the chromosomal DNA,
CC partially digesting with Sau3AI and fractionating on agarose gel.
CC The 3-5 kbp fragment was ligated into pUC18 and E. coli JM109
CC transformed with it. A GT-1 expressing clone was isolated and
CC sequenced. The clone may be used in the development of a drug for
CC dental caries.
XX Sequence 1592 AA;
SQ

CC The present sequence represents a protein from *Streptococcus pneumoniae*.
 CC the nucleic acid sequence encoding the *Streptococcus pneumoniae* protein
 CC can be useful in vaccines for inducing protective antibodies against
 CC *Streptococcus pneumoniae*, for treatment or prevention of infection e.g.
 CC pneumonia, otitis media or meningitis. Probes based on the nucleic acid
 CC are used to detect *Streptococcus pneumoniae* infection (by usual hybridisation or
 CC amplification methods), also for isolating *Streptococcus pneumoniae* genes or their
 CC allelic variants. The protein can be used similarly to detect specific
 CC antibodies in standard immunoassays, especially for diagnosing or
 CC monitoring infections. Antibodies which bind the protein are used to
 CC detect corresponding antigens, to purify the protein and for passive
 CC immunisation (optionally coupled to a toxin). Vaccines are administered,
 CC e.g. by injection, orally or through the skin, typically at 0.01-1000
 CC (especially 10-300) µg/g ml per dose.

XX
 SQ Sequence 591 AA;

Query Match 9.4%; Score 134.5; DB 19; Length 591;
 Best Local Similarity 20.6%; Pred. No. 0.00085;
 Matches 64; Conservative 48; Mismatches 112; Indels 87; Gaps 14;
 QY 2 YNLPQKEVATVRVDRGNVW---KDAIINNLLYVNTIGGGE-----YOKKYGGA 49
 DB 54 YLKSggymaksewvdkafyldqgkmkrnawgtvygaatgkviedwvdsqydw 113
 QY 50 FLDKLOKLYPEFTTKQVSTGVVAIDPSOKIPWSAKYFNGWILHAGSGVVLKADCGQYY 109
 DB 114 FYLKadgqhase-----kew--lqkgkdyfksqyilts---qwi 149
 QY 110 NLGTTTKQPLPIQ--LTGBKQKGNBGFVKGN-----DGNVYFDLAGNWKVNTF 156
 DB 150 nqaynaagakvqqgwlfkqyqsfyfkengnyadkwifenghyfyllksggymaane 209
 QY 157 IEDSGVGNWYFFPDQKQVBNKHFVDVDSYG-----EKGTYFFFLKNGVS 199
 DB 210 lwdk-eswfyfklfdgkmaekewvysqgwyfksqgymtanewlwdkeswfyfksd-- 266
 QY 200 FRGGLVQTD-----NGTYFDNYGKMYRNQFINAGAMIYTLDENGKLIKA-SYNSDAE 251
 DB 267 --gkiaekewvysqgwyfksqgymaknetvdg----yqlgsdgkwlggkttneaa 320
 QY 252 Y-----PTSTDV 258
 DB 321 Yqygvptanv 331

RESULT 5
 AAY81517
 ID AAY81517 standard; Protein; 659 AA.
 AC AAY81517;
 XX
 XX 24-MAY-2000 (first entry)
 DE Streptococcus pneumoniae type 4 protein sequence #17.

XX Streptococcus pneumoniae; vaccine; screening; protein antigen;
 KW antibacterial; antiinflammatory; meningitis; infection; diagnosis;
 KW pneumococcal disease.

XX Streptococcus pneumoniae.

XX WO200006737-A2.

XX 10-FEB-2000.

XX 27-JUL-1999; 99WO-GB02451.

XX 27-JUL-1998; 98GB-0016337.

XX 19-MAR-1999; 99US-0125164.

PA (MICR-) MICROBIAL TECHNIQS LTD.

XX gilbert CFG, Hansbro PM;
 PI WPI; 2000-195300/17.
 XX New Streptococcal protein, useful as a vaccine, for diagnosis of
 PT pneumococcal diseases and for screening agents capable of antagonizing
 PT or inhibiting expression of the protein
 XX Claim 1; Page 74; 108pp; English.
 XX AAY81501 to AAY81679 represent specifically claimed protein sequences
 CC isolated from *Streptococcus pneumoniae* AAY05407 to AAY05590 represent
 CC specifically claimed nucleotide sequences isolated from *S. pneumoniae*.
 CC The sequences have antibacterial and antiinflammatory properties.
 CC The protein sequences, and fragments of them, are useful as immunogens
 CC and/or antigens. The nucleotide sequences can be used in vaccines and
 CC diagnostic assays. The proteins and nucleotides can be useful for the
 CC detection and diagnosis of *S. pneumoniae*. The protein sequences are also
 CC useful for screening an agent capable of antagonising, inhibiting or
 CC interfering with the function or expression of the proteins in which the
 CC agent is useful for treatment or prophylaxis of *S. pneumoniae* infection
 CC and meningitis. AAY05591 to AAY05614 represent primers used in the
 CC exemplification of the present invention.

XX Sequence 659 AA;

Query Match 9.4%; Score 134; DB 21; Length 659;
 Best Local Similarity 23.2%; Pred. No. 0.0011;
 Matches 60; Conservative 41; Mismatches 108; Indels 50; Gaps 13;
 QY 37 IGGGEYQKYGAFIDK--LQKLYPEFTTKQVSTGVVAIDPSQKITEH-----SAKYFNG 89
 DB 153 lkgkdyfksqyiltsqinqayvnasgk-vqggwlfkdkyqsfyfkengnyadk 209
 QY 90 TNLHRSQGVYVLKADGGQYVNLGTTTKQ-FLPIQLTGSKQGNBGFVKGNDGNYFYVDLA 148
 DB 210 ewifenghyfyllksggymaaneftwksfwylkfdg-kmaekewvysqgwyfksq 268
 QY 149 GNMVKNTPYIESVGNWYFFPDQKQVBNKHFVDVDSYG-----EKGTY 191
 DB 269 gyntanewlwdk-eswfyfksdgkiaekewvysqgwyfksqgymtanewlwdkesw 327
 QY 192 FFLKNGVSFRGLVQTD-----NGTYFDNYGKMYRNQFINAGAMIYTLDENGKLIKA 244
 DB 328 fylksd----gkiaekewvysqgwyfksqgymaknetvdg----yqlgsdgkwlgg 379
 QY 245 -SYNSDAEY-----PTSTDV 258
 DB 380 kttnenaayqygvptanv 398

RESULT 6
 AAY81660
 ID AAY81660 standard; Protein; 319 AA.

XX AAY81660;

XX 24-MAY-2000 (first entry)

XX Streptococcus pneumoniae protein sequence ID309.

XX Streptococcus pneumoniae; vaccine; screening; protein antigen;
 KW antibacterial; antiinflammatory; meningitis; infection; diagnosis;
 KW pneumococcal disease.

XX Streptococcus pneumoniae.

XX WO200006737-A2.

XX 10-FEB-2000.

PF 27-JUL-1999; 99WO-GB02451.
 XX
 PR 27-JUL-1998; 98GB-0016337.
 PR 19-MAR-1999; 99US-0125164.
 XX
 PA (MICR-) MICROBIAL TECHNIKS LTD.
 XX
 PI Gilbert CFG, Hansbro PM;
 XX
 DR WPI; 2000-195300/17.
 XX

PT New Streptococcal protein, useful as a vaccine, for diagnosis of
 PT pneumococcal diseases and for screening agents capable of antagonizing
 PT or inhibiting expression of the protein
 XX

PS Claim 2; Page 98; 108pp; English.
 XX

CC AAY81501 to AAY81679 represent specifically claimed protein sequences
 CC isolated from Streptococcus pneumoniae. AAA05407 to AAA05590 represent
 CC specifically claimed nucleotide sequences isolated from S. pneumoniae.
 CC The sequences have antibacterial and antiinflammatory properties.
 CC The protein sequences, and fragments of them, are useful as immunogens
 CC and/or antigens. The nucleotide sequences can be used in vaccines and in
 CC diagnostic assays. The proteins and nucleotides can be useful for the
 CC detection and diagnosis of S. pneumoniae. The protein sequences are also
 CC useful for screening an agent capable of antagonising, inhibiting or
 CC interfering with the function or expression of the proteins in which the
 CC agent is useful for treatment or prophylaxis of S. pneumoniae infection
 CC and meningitis. AAA05591 to AAA05614 represent primers used in the
 CC exemplification of the present invention.
 XX

SQ Sequence 319 AA;

Query Match 9.2%; Score 132; DB 21; Length 319;
 Best Local Similarity 21.9%; Pred. No. 0.00062;
 Matches 62; Conservative 44; Mismatches 99; Indels 78; Gaps 14;

QY 8 EVATVTRVDDRCNWKDALINNLYVWNTIGGEGYQKYGGA----- 49
 Db 48 elgfshlvngclmqvdpvngaw---dvsggwnaetyaavellshstkeefmtdyrl 104
 QY 50 FLDKLQKLYPEITFTKKQVSTGVA-----IDPSQKITEW--SAKYF 87
 Db 105 yieilrnladeaglktdtsglakitheyctnqpnhsdhvpyylakwlsreql 164
 QY 88 -----NCTNLRGSGYVLKADCGGYNLGTTTKQFLPIQLTGCKKQNEGFPVKGNDGNY 142
 Db 165 kadienglti---etgw-qkndtgyw-----vhsdgsypkdkfekln-gtw 206
 QY 143 FYFDLAGNNMKNFTFEDSGNMYFFDQDGKMYEN-KHFVDVDSYGEKGTYPFLKNGVSFR 201
 Db 207 yfidsagymldrwrkhdgwnywfndsgematgwkkiaad-----kwyfneeg-amk 258
 QY 202 GGLVTDNTGTYFD-NYGMVNRNPTINA--GAMIYTLDENCKL 241
 Db 259 tgwkykdtwyldakegamsnafigsadgtgywylkpdgtl 301

RESULT 7

AAW81667
 ID AAY81667 standard; Protein; 678 AA.
 XX
 AC AAY81667;
 XX

DT 24-MAY-2000 (first entry)

DE Streptococcus pneumoniae protein sequence ID204 - 4117.1.

KW Streptococcus pneumoniae; vaccine; screening; protein antigen;
 KW antibacterial; antiinflammatory; meningitis; infection; diagnosis;
 KW pneumococcal disease.
 XX

OS Streptococcus pneumoniae.

XX WO200006737-A2.

XX 10-FEB-2000.

XX 27-JUL-1999; 99WO-GB02451.

XX 27-JUL-1998; 98GB-0016337.

XX 19-MAR-1999; 99US-0125164.

XX (MICR-) MICROBIAL TECHNIKS LTD.

XX Gilbert CFG, Hansbro PM;

XX WPI; 2000-195300/17.

XX New Streptococcal protein, useful as a vaccine, for diagnosis of

PT pneumococcal diseases and for screening agents capable of antagonizing

PT or inhibiting expression of the protein

XX Claim 6; Page 99; 108pp; English.

CC AAY81501 to AAY81679 represent specifically claimed protein sequences
 CC isolated from Streptococcus pneumoniae. AAA05407 to AAA05590 represent
 CC specifically claimed nucleotide sequences isolated from S. pneumoniae.
 CC The sequences have antibacterial and antiinflammatory properties.

CC The protein sequences, and fragments of them, are useful as immunogens
 CC and/or antigens. The nucleotide sequences can be used in vaccines and in
 CC diagnostic assays. The proteins and nucleotides can be useful for the
 CC detection and diagnosis of S. pneumoniae. The protein sequences are also
 CC useful for screening an agent capable of antagonising, inhibiting or
 CC interfering with the function or expression of the proteins in which the
 CC agent is useful for treatment or prophylaxis of S. pneumoniae infection
 CC and meningitis. AAA05591 to AAA05614 represent primers used in the
 CC exemplification of the present invention.
 XX

SQ Sequence 678 AA;

Query Match 9.0%; Score 129; DB 21; Length 678;

Best Local Similarity 24.4%; Pred. No. 0.0033;

Matches 62; Conservative 30; Mismatches 96; Indels 66; Gaps 13;

QY 19 GNWVKDALINNLYVWNTIGGEGYQKYGGAFLKQLKLYPEITFTKKQVSTGVAIDPSQX 78

Db 4 gefwnklnkr-----grlmkkrvfilall-----fflaspegamasdgt-- 46

QY 79 ITEWSAKYFNGTNIHRGSGYVLKADGSGQYVN--LGTTTKQFLPIQLTGCKKQNEGFFV 135

Db 47 ---wggkqy-----lkedgsqaanewvfdthgyswfyikadanyaenewlk 89

QY 136 KGNCGNYFYDLAGNMVKNFTIEDSVGNWYFFDQDGKMYENKHFVDVDSYGEKG----- 189

Db 90 qgdd--yfyksggymaksewedk-gafyyldqdgkmkrna-wvgtisyvgatgkvied 145

QY 190 -----TYEFLK-NGVSPRGLVQTDNGTYIYFDNYGKMYNRNPTIN-----AGAMIYT 234

Db 146 wvdsqydawfyikadqgkhaekewliqkgdyfksqgylltsqwhqayvnasgkv-- 203

QY 235 EDENGLIKFASVNS 248

Db 204 --qggwlfkdyqs 215

RESULT 8

AAW61214
 ID AAW61214 standard; Protein; 569 AA.
 XX

AC AAW61214;

XX 02-OCT-1998 (first entry)

DT

XX

DE Streptococcus pneumoniae SP0089 protein.
KW Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;
KW detection; pneumonia; otitis media; meningitis.
XX
OS Streptococcus pneumoniae.
XX
PN W09818930-A2.
XX
PD
XX
XX 07-MAY-1998.
XX
XX 30-OCT-1997; 97MO-US19422.
XX
XX 31-OCT-1996; 96US-0029960.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Choi GH, Hromockyj A, Johnson LS, Kunsch CA;
XX WPI, 1998-272224/24.
XX N-FSDB; AAV27400.
XX
XX Nucleic acid encoding antigenic peptide(s) from Streptococcus
PT pneumoniae - or their epitope-containing fragments, useful in
PT protective or therapeutic vaccines, and for diagnosis
XX
XX Claim 11; Page 80; 118pp; English.
XX
XX The present sequence represents a protein from Streptococcus pneumoniae.
XX The nucleic acid sequence encoding the Streptococcus pneumoniae protein
CC can be useful in vaccines for inducing protective antibodies against
CC Streptococcus pneumoniae, for treatment or prevention of infection e.g.
CC pneumonia, otitis media or meningitis. Probes based on the nucleic acid
CC are used to detect Streptococcus infection (by usual hybridisation or
CC amplification methods), also for isolating Streptococcus genes or their
CC allelic variants. The protein can be used similarly to detect specific
CC antibodies in standard immunoassays, especially for diagnosing or
CC monitoring infections. Antibodies which bind the protein are used to
CC detect corresponding antigens, to purify the protein and for passive
CC immunisation (optionally coupled to a toxin). Vaccines are administered,
CC e.g. by injection, orally or through the skin, typically at 0.01-1000
XX (especially 10-300) mu g/ml per dose.
XX
SQ Sequence 569 AA;

Query Match 9.0%; Score 128; DB 19; Length 569;
Best Local Similarity 22.8%; Pred. No. 0.0032;
Matches 59; Conservative 41; Mismatches 109; Indels 50; Gaps 13;

QY 37 IGGGEYQKYGAFDOK--LQKLYPEIFTKKQVSGVAIDPSOKITEW-----SAKYFNG 89
DB 69 lKqkdyfksqgylltsqinqayvnaagak-vqggwlfdd--kqyswfykengnyadk 125
QY 90 TWLHGGGYVLKADGGYNYLGTGTTKQ-FLPIQLTGKKGNEGFVKGNQNYFYFDIA 148
DB 126 ewifenghyvylksggymaanewlwkdeswfyllkfdg-kmaekewydsqawyyfksq 184
QY 149 GNMVKNFTFIEDSGHWYFDDQKMKVENKHFVDVDSYG-----EKGTY 191
DB 185 gymtanewlwdk-eswfyllksdgkkaekewydsqawyyfksqgyntanewlwdkesw 243
QY 192 FFLKNGVSPRGLVQTD-----NGTYFDYHGKWRNQTINAGAMITYTLDENGKLIK 244
DB 244 fylksd----gkiaekewydsqawyyfksqgygmaknetvdg----yqigsdskwlgg 295
QY 245 -SYNSDAEY-----PTSTDV 258
DB 296 kttnenaayqvvpvtanv 314

RESULT 9
AAB98217

ID AAB98217 standard; Protein; 214 AA.
XX
AC AAB98217;
XX
DT 20-AUG-2001 (first entry)
XX
DE Peptostreptococcus micros Fiba protein sequence.
XX
KW Human; hfiba; Fiba; fibril-like surface structure.
XX
XX Peptostreptococcus micros.
OS
XX CN1281045-A.
PN
PD 24-JAN-2001.
XX
XX 27-JUN-2000; 2000CN-0116792.
PF
XX 27-JUN-2000; 2000CN-0116792.
PR
XX (NANP-) NANPANG RES CENT STATE HUMAN GENE GROUP.
PA
XX Yang Y, Xiao H, Kang B;
PI WPI; 2001-282655/30.
XX
XX Human fibrous surface structure subunit protein and its coding sequence
PT
PT
XX
XX Example 2; Fig 2; 19pp; Chinese.
XX
CC The present invention describes a human fibril-like surface structure
CC subunit protein, designated hfiba. Also described are methods for the
CC preparation and detection of hfiba protein and nucleotide sequences.
CC The present sequence represents the Peptostreptococcus micros Fiba
CC protein which is given in comparison with the hfiba protein sequence
CC in an example from the present invention.
XX
XX Sequence 214 AA;

Query Match 8.8%; Score 126; DB 22; Length 214;
Best Local Similarity 21.3%; Pred. No. 0.0013;
Matches 58; Conservative 44; Mismatches 70; Indels 100; Gaps 15;

QY 11 TVTRVDD---RCNFWK-----DAIINNLXYVNTTIGGGEYQKYGAF 50
DB 4 tlpkfdvkvvggtwkfitwtvknkvepienealvenedlrliq----gwk-----f 53
QY 51 L-DKLOKLYPEIFTKKQVSTGV---AIDPSOKITEWSAKYFNGTNILHRSGGYVLKADG 105
DB 54 vkdgqepnptkykwqnedgtwtqstgkkekikd-swkin-----newryfdkdg 105
QY 106 -----GQYNL---GTTTKQFLPIQLTGKKGNEGFVKGNQNYFYFDLAGNM 151
DB 106 kmianswfeengkwvyleasgmsk-----newvyk--dgnwyvyanasgri 149
QY 152 VKNFIEDSVGNWYFFDDQKMKVENKHFVDVDSYCEKTYFFLKNQVSPRGLVQTDNGT 211
DB 150 sqnewlvd-gnwvyvnanasgriaanewfm-----yvgk-----w 182
QY 212 YFFNYGKWRNQTINAGAMITYTLDENGKLIK 243
DB 183 yyaesdgriaqgktlrinnvyytfdngvwir 214

RESULT 10
AAY81659
ID AAY81659 standard; Protein; 628 AA.
XX
AC AAY81659;
XX
DT 24-MAY-2000 (first entry)

XX Streptococcus pneumoniae protein sequence ID308.
DE Streptococcus pneumoniae; vaccine; screening; protein antigen;
KW antibacterial; antinflammatory; meningitis; infection; diagnosis;
KW pneumococcal disease.
XX Streptococcus pneumoniae.
OS Streptococcus pneumoniae.
XX W0200006737-A2.
XX 10-FEB-2000.
XX 27-JUL-1999; 99WO-GB02451.
XX 27-JUL-1998; 98GB-0016337.
XX 19-MAR-1999; 99US-0125164.
XX (MCR-) MICROBIAL TECHNIQS LTD.
XX Gilbert CFG, Hansbro PM;
PI WPI: 2000-195300/17.
XX New Streptococcal protein, useful as a vaccine, for diagnosis of
PT pneumococcal diseases and for screening agents capable of antagonizing
PT or inhibiting expression of the protein
XX Claim 2: Page 98; 108pp; English.
XX AAY81501 to AAY81679 represent specifically claimed protein sequences
CC isolated from Streptococcus pneumoniae. AAY81501 to AAY81590 represent
CC specifically claimed nucleotide sequences isolated from S. pneumoniae.
CC The sequences have antibacterial and antiinflammatory properties.
CC The protein sequences, and fragments of them, are useful as immunogens
CC and/or antigens. The nucleotide sequences can be used in vaccines and in
CC diagnostic assays. The proteins and nucleotides can be useful for the
CC detection and diagnosis of S. pneumoniae. The protein sequences are also
CC useful for screening an agent capable of antagonising, inhibiting or
CC interfering with the function or expression of the proteins in which the
CC agent is useful for treatment or prophylaxis of S. pneumoniae infection
CC and meningitis. AAY81591 to AAY81614 represent primers used in the
CC exemplification of the present invention.
XX Sequence 628 AA;
Query Match 8.7%; Score 125; DB 21; Length 628;
Best Local Similarity 20.9%; Pred. No. 0.0069;
Matches 71; Conservative 58; Mismatches 104; Indels 106; Gaps 20;
QY 21 VWKDAIINN-LYVWNTIG-----GGEY-----QKXGGAF----- 50
DB 204 lwd-----nsnslisvkvngkkyilggldnhvhgaedkygpllgkvdlnkfnhhdtnks 260
QY 51 --LQKLYLPEIFTKQVS---TGVAIDPSQKITENSAKYFNCTNLHRSGSVLKA 104
DB 261 ntkdfiknlspilvgtsslpwngvdsyvnwklgerinaasakdydvtdirkd 320
QY 105 GQYVYVNLGTYNQFLPIQI-----TGEXKQG-NEGFKVKGNDGNYFYFD 146
DB 321 g--fvnistsyxpipsfsgwhksaygnwygapdstgeyavgwne-----legewyyfn 373
QY 147 LAGNVAKNY-----FIEDS-----VGNWYFFDQDGM-----VENK---HFV 180
DB 374 qtglllqkqkwnhwltyltdsgasaknwkiagiwyfinkendmeigwqkeqwyyl 433
QY 181 DVDSYGEK-----TYFFELKNGVSFEGGLVQTDNGYFYFDNKGKMYRNQNTINAGMIY 233
DB 434 dvdgamktgvlqmvgwqwyfapsgeamkwvdketwyndstgvm-ktgeievagqhy 491
QY 234 TLDENGKLIKASYNDAE-----YPT--STDVCKMIDQNK 266
| : : : | : : | | | | | : : |

DB 492 yledsgam-kqgwhkkandwyfyktdgsgsavgwikdkd 529
RESULT 11
AAR95016
ID AAR95016 standard; Protein; 2710 AA.
XX AAR95016;
AC AAR95016;
XX 08-JUL-1996 (first entry)
XX C. difficile toxin A.
DE C. difficile toxin A.
XX Toxin A; cytotoxin; enterotoxin; fusion protein; antitoxin;
KW vaccine; diarrhoea; diagnosis; therapy.
XX Clostridium difficile VPI strain 10463 (ATCC 10463).
OS Clostridium difficile VPI strain 10463 (ATCC 10463).
XX W09612802-A1.
PN W09612802-A1.
XX 02-MAY-1996.
PD 02-MAY-1996.
XX 23-OCT-1995; 95WO-US13737.
XX 07-JUN-1995; 95US-0480604.
PR 24-OCT-1994; 94US-0329154.
PR 16-MAR-1995; 95US-0405496.
PR 14-APR-1995; 95US-0422711.
XX (OPHI-) OPHIDIAN PHARM INC.
XX Firca JR, Kink JA, Padhye NV, Stafford DC, Thalley BS;
PI Williams JA;
XX WPI: 1996-230603/23.
DR N-PSDB; AAT29248.
XX Fusion proteins comprising non-toxin protein and part of toxin
PT useful to form anti-toxins against Clostridium botulinum type A, and
PT C. difficile type toxins, and to treat C. difficile intoxication,
PT partic. diarrhoea
XX Claim 63; Page 290-302; 434pp; English.
XX Clostridium difficile VPI strain 10463 toxin A (AAR95016), the
CC product of the toxin A gene (AAR29248), is a potent cytotoxin that
CC plays a direct role in damaging gastrointestinal tissues and is
CC associated with diarrhoeic disease. It can be obtd. by expression in
CC transformed E. coli hosts of portions of DNA that together cover the
CC entire toxin A gene. Toxin A, and portions of it (see also
CC AAR95014-15 and AAR95017), pref. expressed as fusions to polyhistidine
CC affinity tags or maltose binding protein, are used to raise avian
CC antibodies useful as antitoxins or diagnostics, and in vaccine prodn.
XX Sequence 2710 AA;
Query Match 8.7%; Score 125; DB 17; Length 2710;
Best Local Similarity 24.1%; Pred. No. 0.052;
Matches 60; Conservative 37; Mismatches 101; Indels 76; Gaps 17;
QY 16 DDRGNVWKDAI-INNNLYVWN-----TIGGEYQKKGAGFLDKLQKLYPEIFT 63
DB 1824 dedaklvkglinlnnsifypdplefnlvtgwtng---kky---yfd----- 1865
QY 64 KKQYSTGAIDPSQKITENSAKYFNCTNLHRSGSVYLKA-DGQYVYVNLGTYNQFLP-- 120
DB 1866 ---intgaal-tsykiingkhfyfndngvmqlg---vfkpgdgyefapantqnnlegq 1918
QY 121 -----IQLTGEEK--QGNFVKG-----NDGNVYFYDLAGNKNVNTFIEDSVGNWYF 166
DB 1919 aivgskfltingtkkyfyfdnnskavtwtinnekyyfn--pnnaiaavglqvadnky 1976
| : : : | : : | | | | | : : |

```
OY 167 FDDGGRKWE-----NKHFVDVDS---YGEK---GTFFFLKNGVSPRGLVQTDNGT 211
Db 1977 fnpdtailskgwqvtngsryyfdtdtaiafngyktidgkhfydsdcvkvlgvfstngf 2036

OY 212 YFF---DNYGKVMRNQTNAGAMIYTLDENGKLKASYNDSA 250
Db 2037 eyfapantynnleqgaivqskfcti--ngkkyfyfannska 2076

RESULT 12
AAW68387
ID AAW68387 standard; Protein: 2710 AA.
XX AC AAW68387;
XX DT 07-DEC-1998 (first entry)
XX DE Clostridium difficile toxin A.
XX KW Antibiotin; vaccine; cytotoxin; toxin A; intoxication; immunogen;
XX KW pseudomembranous enterocolitis.
XX OS Clostridium difficile.
XX PN W09808540-A1.
XX PD 05-MAR-1998.
XX PF 28-AUG-1997; 97WO-US15394.
XX PR 28-AUG-1996; 96US-0704159.
XX PA (OPH1-) OPHIDIAN PHARM INC.
XX PT Thalley BS, Williams JA;
XX DR WPI; 1998-230234/20.
XX DR N-PSDB; AAV30560.
XX Host cell containing recombinant expression vector encoding
XX Clostridium botulinum type B or E toxin - useful to treat humans
XX and other animals at risk of intoxication with clostridial toxin
XX Example 15; Page 220-230; 428pp; English.
XX This is the amino acid sequence of Clostridium difficile toxin A,
XX deduced from the coding region (see AAV30560) of the toxin A gene.
XX Toxin A is a potent cytotoxin that plays a direct role in damaging
XX gastrointestinal tissues. Severe cases of C. difficile
XX intoxication result in pseudomembranous colitis. This would be
XX prevented by neutralising the effects of toxin A in the
XX gastrointestinal tract. Examples are provided of the production
XX of recombinant C. difficile toxin A in host cells and of the in
XX vivo neutralisation of toxin A by antibodies against recombinant
XX toxin A polypeptides. The invention specifically relates to
XX recombinant proteins derived from Clostridium botulinum toxins
XX (see AAW68389-400) and their use as immunogens for the production of
XX vaccines and antitoxins.
XX Sequence 2710 AA;

Query Match 8.7%; Score 125; DB 19; Length 2710;
Best Local Similarity 24.1%; Pred. No. 0.052;
Matches 68; Conservative 37; Mismatches 101; Indels 76; Gaps 17;

OY 16 DGRGMWKKDAI-INNNLYVWN-----TJGGGYOKKYGGAFUDKQLKLYPEIFT 63
Db 1824 dedsklvkglnimnslfyfpiefnlvtgwtng---kky-----yfd----- 1865

OY 64 KQVGTGVAIDPSKITEASRYFNCTNLRHSGGYVLA-DGGQYVNLGTTTKQFLP-- 120
Db 1866 ---intgaal-tsykiingkhfyfandgvmlq---vfkpgpdyefapantqnnieqg 1918
```

```
OY 121 -----IQLTGEKK--QNEGFKVG-----NDGNFYFYDLAGNMVKNFTFIEDSVGNWYF 166
Db 1919 aivvyskfitngkkyfyfddnskaavgtrifionekyyfn-pnnalaaavglgvidnnkyy 1976

OY 167 FDDGGRKWE-----NKHFVDVDS---YGEK---GTFFFLKNGVSPRGLVQTDNGT 211
Db 1977 fnpdtailskgwqvtngsryyfdtdtaiafngyktidgkhfydsdcvkvlgvfstngf 2036

OY 212 YFF---DNYGKVMRNQTNAGAMIYTLDENGKLKASYNDSA 250
Db 2037 eyfapantynnleqgaivqskfcti--ngkkyfyfannska 2076

RESULT 13
AAW06593
ID AAW06593 standard; Protein: 453 AA.
XX AC AAW06593;
XX DT 26-OCT-1999 (first entry)
XX DE CLYTA-MAGE-3-His fusion protein.
XX KW MAGE-3; CLYTA-MAGE-3-His; fusion protein; tumour; melanoma;
XX KW breast cancer; bladder cancer; lung cancer; colon cancer;
XX KW head and squamous cell carcinoma; oesophagus carcinoma; vaccine;
XX KW human.
XX OS Chimeric - Streptococcus pneumoniae.
XX OS Chimeric - Homo sapiens.
XX OS Chimeric - synthetic.
XX FH Key Location/Qualifiers
XX Misc-difference 245 /note= "encoded by AAG"
XX Misc-difference 286 /note= "encoded by GAT"
XX Misc-difference 305 /note= "encoded by GTC"
XX Misc-difference 329 /note= "encoded by ACA"
XX Misc-difference 331 /note= "encoded by TTC"
XX Misc-difference 335 /note= "encoded by ATC"
XX Misc-difference 341 /note= "encoded by AAA"
XX Misc-difference 369 /note= "encoded by TTC"
XX Misc-difference 379 /note= "encoded by TAT"
XX Misc-difference 409 /note= "encoded by ATT"
XX Misc-difference 428 /note= "encoded by CGC"
XX Misc-difference 433 /note= "encoded by CTC"
XX Misc-difference 438 /note= "encoded by GCT"
XX PN W09940188-A2.
XX PD 12-AUG-1999.
XX PF 02-FEB-1999; 99WO-EP00560.
XX PR 06-FEB-1998; 98GB-0002650.
XX PR 05-FEB-1998; 98GB-0002543.
XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX Cabezon Silva T, Cohen J, Slaoui MM, Vinals Bassols C;
```

```

XX WPI; 1999-494293/41.
DR N-PSDB; AAX87596.
XX
XX New protein derivatives used in cancer vaccine therapy for treating
PT a range of cancers including melanomas, carcinomas and cancers of
PT breast
XX
XX Example 10; Page 71-72; 72pp; English.
XX
XX The present sequence represents a fusion protein composed of
CC the C-terminal portion of the Streptococcus pneumoniae LYTA
CC protein (CLYTA), the human MAGE-3 tumour-associated antigen and a
CC hexahistidine tail. A vector designed for recombinant expression
CC of the fusion protein in Escherichia coli is provided. The
CC CLYTA moiety provides expression of soluble fusion protein,
CC facilitates affinity purification, and also acts as a T-helper
CC epitope. The invention relates to MAGE proteins fused to an
CC immunological fusion partner, e.g. CLYTA-MAGE-3-His. These novel
CC fusion proteins provide vaccines for immunotherapy of melanomas or
CC other MAGE-associated tumours like breast, bladder, lung and
CC non-small cell lung cancer, head and squamous cell carcinoma, colon
CC carcinoma and oesophagus carcinoma.
XX
XX Sequence 453 AA;
SQ
Query Match 8.4%; Score 120.5; DB 20; Length 453;
Best Local Similarity 29.0%; Pred. No. 0.012;
Matches 38; Conservative 24; Mismatches 46; Indels 23; Gaps 7;
QY 139 DGNYYFYDLGNMKNTEIYDSVGNWYFDDQKAVEN-KHFVDVDSYGEKGTFFELKNG 197
DB 21 ngtwyyfssgymiadrrwrkntdgnwyyfwnsgematgwkxiad-----kwyfneeg 73
QY 198 YSFRGLYOTDNGTYTFD-NTCKMVRNOTINA--CAMTYTLDENGLIKASVNSDAEYPT 254
DB 74 -amktgwkykdtwylldakegamvsnafiqsadtgwylykpdgtl-----adrp- 123
QY 255 STDVCKMLDQN 265
DB 124 --elasmidmd 132
RESULT 14
AAY43219
ID AAY43219 standard; Protein; 606 AA.
XX
XX AAY43219;
XX
XX 13-JAN-2000 (first entry)
XX
XX E. rhusiopathiae surface layer protein sequence.
XX
XX Surface layer protein; immunological detection; antigen; pig erysipelas;
XX genetic detection.
XX
XX Erysipelothrix rhusiopathiae.
XX
XX JP11262391-A.
XX
XX 28-SEP-1999.
XX
XX 17-MAR-1998; 98JP-0067258.
XX
XX 17-MAR-1998; 98JP-0067258.
XX (CHCC ) CHISSO CORP.
XX
XX WPI; 1999-603802/52.
XX N-PSDB; AAZ31581.
XX
XX New surface layer protein of Erysipelothrix rhusiopathiae - useful for
PT

```

```

PT detecting antibody against E. rhusiopathiae
XX
XX Claim 1; Page 8-11; 14pp; Japanese.
XX
XX This sequence is the Erysipelothrix rhusiopathiae surface layer
CC protein of the invention. The surface layer protein can be used as an
CC antigen in a method for the immunological detection of E. rhusiopathiae.
CC It can also be used as an antigen: (1) in a method for the immunological
CC detection of an antibody against E. rhusiopathiae; and (2) in a method
CC for the immunological detection of pig erysipelas. The DNA sequence can
CC be used in a method for the genetic detection of E. rhusiopathiae or pig
CC erysipelas. The methods are simple and rapid.
XX
XX Sequence 606 AA;
SQ
Query Match 8.4%; Score 120.5; DB 20; Length 606;
Best Local Similarity 23.6%; Pred. No. 0.017;
Matches 39; Conservative 30; Mismatches 59; Indels 37; Gaps 8;
QY 120 PIQLTGKKGQNECFVKNGDGNYYFYDLGNMKNTEIYDSVGNWYFDDQD-----KM 173
DB 435 pydskpseskgwik-kdnkwyfleksgma--tgwkvadkwylydntgalvtgwk 491
QY 174 VENKHFVDVDSYGEKGT-----YFELKNGYSFRGLVQTDNGTYTFDNYG-----K 219
DB 492 vankwyy-leksgamatgwkvsnkwyylensgamatgwkvsnkwyylensgamatgwk 550
QY 220 MVRNQTIT---NAGAMI-----YTLDENGKLKIKASVNSDAE 251
DB 551 kvsnkwyylensgamatgwkvsnkwyylldksgmmvtgsksidgk 595
RESULT 15
AAY25380
ID AAY25380 standard; Protein; 239 AA.
XX
XX AAY25380;
XX
XX 06-SEP-1999 (first entry)
XX
XX HPV fusion protein CLYTA-E7-His/HPV16.
XX
XX Fusion protein; E6 protein; E7 protein; E6/E7; immunomodulator; tumour;
XX immunological fusion partner; CpG oligonucleotide; immune response;
XX HPV antigen; prevention; treatment.
XX
XX Synthetic.
XX OS Human papillomavirus.
XX
XX WO9933868-A2.
XX
XX 08-JUL-1999.
XX
XX 18-DEC-1998; 98WO-EP08563.
XX
XX 24-DEC-1997; 97GB-0027262.
XX
XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
XX Dalemans WLJ, Gerard CMG;
XX
XX WPI; 1999-405485/34.
XX N-PSDB; AAX78796.
XX
XX Composition comprising an E6, E7 or E6/E7 fusion protein from HPV to
XX induce immune response to HPV
XX
XX Example VII; Page 53; 62pp; English.
XX
XX AAX78791-X78801 represent nucleic acid sequences which encode novel
XX constructs comprising an E6 or E7 protein or E6/E7 fusion protein from
XX HPV (represented in AAY25375-Y25386). These constructs are optionally
XX

```


[illegible]


```

FT Peptide 1..37
FT /label= Sig_peptide
FT Protein 38..929
FT /label= Mat_protein
FT Domain 38..637
FT /label= Alpha-helix
FT Region 69..637
FT /label= Coiled-coil
FT /note= "breaks in 7-residue periodicity of the
FT coiled-coil occur at amino acids 136-141,
FT 261-304 and 383-387"
FT
FT Region 41..242
FT /label= Repeat_1
FT /note= "alpha-helical repeat region"
FT Region 332..492
FT /label= Repeat_2
FT /note= "alpha-helical repeat region"
FT Domain 627..689
FT /label= Proline-rich
FT Domain 913..929
FT /label= C-terminal
FT
XX W09709994-A1.
XX
XX 20-MAR-1997.
XX
XX 16-SEP-1996; 96WO-US14819.
XX
XX 15-SEP-1995; 95US-0529055.
XX
XX (UABR-) UAB RES FOUND.
XX
XX Briles DE, Brooks-Walter A, Crain MJ, Hollingshead S;
XX McDaniel LS, Swiatlo E, Tart R, Yother J;
XX WPI; 1997-202002/18.
XX N-PSDB; AAT61728.
XX
XX Streptococcus pneumoniae surface protein pspC and truncated PspA -
XX used in vaccines for protecting animals against S.pneumoniae
XX infection
XX
XX Disclosure; Fig 13; 296pp; English.
XX
XX This sequence comprises the pneumococcal protein surface C (pspC)
XX of Streptococcus pneumoniae strain EF6796. The sequence was
XX deduced from the pspC gene (AAT61728). Like PspA, PspC has an
XX alpha-helical coiled-coil region, proline-rich central region,
XX repeat regions, with a choline binding motif, and a C-terminal
XX 17-amino acid tail. The 2 polypeptides share 3 regions of high
XX sequence identity. One is a protection-eliciting region present
XX within the alpha-helical domain. The others are the proline-rich
XX domain and a repeat domain shared with other choline-binding
XX proteins and thought to play a role in cell surface association.
XX PspC and PspA polypeptides, and their fragments, can be used in
XX vaccines to protect against S. pneumoniae infection and hence for
XX the prevention of diseases such as otitis media, meningitis,
XX bacteraemia and pneumonia.
XX
XX Sequence 929 AA;

```

```

Query Match 8.2%; Score 117.5; DB 18; Length 929;
Best Local Similarity 22.2%; Pred. No. 0.059;
Matches 43; Conservative 34; Mismatches 80; Indels 37; Gaps 8;

QY 75 PSQKTTESAK-----YFNGTH-----ILHRGSGYVLKADG-----GQYVNL-- 111
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
686 petpktgwkqngmwyfntdgsmtgwlqngswyylnsngamatgwlqngswyylns 745
QY 112 -GTTTKQQLPQLGKKGQNGFVKG---NDGNYFYDLAGNMVKNFFIEDSVGNMYFF 167
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
746 ngamatgwlqngswyylngangdmatgwlqngswyylngangdmatgwfqyn--gswyyl 803

```

```

QY 168 DQDGKAVENKHFVDVDSYSGEKGTYFFLKNGVSFRRGLVQFDNGTYFYFDNYGKAVRNQTTN 227
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
804 nangdma-----tgwfqngswyylnsngamatgwlqngswyylnsngamvtgwlqp 856
QY 228 AGAMIYTLDENGKL 241
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
857 ngswyylngangsm 869

RESULT 23
AAW87971
ID AAW87971 standard; peptide; 862 AA.
XX
XX AAW87971;
XX
XX 29-APR-1999 (first entry)
XX
XX Toxin A immunogenic fragment p9/10.
XX
XX Toxin A; immunogenic fragment; vaccine; diagnosis; infection.
XX
XX Clostridium difficile.
XX
XX W09859053-A1.
XX
XX 30-DEC-1998.
XX
XX 19-JUN-1998; 98WO-GH01805.
XX
XX 07-JAN-1998; 98GB-0000321.
XX
XX 20-JUN-1997; 97GB-0013146.
XX
XX (UNLO ) IMPERIAL COLLEGE SCI TECHNOLOGY & MED.
XX (QUEE-) QUEEN MARY & WESTFIELD COLLEGE.
XX
XX Douce G, Dougan G, Ward SJ, Wren BW;
XX WPI; 1999-081281/07.
XX
XX New Isolated Clostridium difficile proteins - comprise fragments of
XX toxin A, used to develop agents for prophylaxis, treatment or
XX diagnosis of C. difficile infections
XX
XX Example 1; Fig 8; 82pp; English.
XX
XX This sequence represents an immunogenic fragment of Clostridium difficile
XX (CD) Toxin A, designated fragment p9/10. This sequence is encoded by
XX nucleotides 5530-8115 of the Toxin A gene. The invention relates to
XX fragments of toxin A of CD which are non-toxic and immunogenic. They can
XX be used in the preparation of vaccines against CD infection. The amino
XX acid molecules and antibodies against them can be used in the preparation
XX of an agent for the prophylaxis or treatment of a CD infection. They can
XX also be used for detection and diagnosis of CD infection.
XX
XX Sequence 862 AA;

Query Match 8.2%; Score 117; DB 20; Length 862;
Best Local Similarity 24.4%; Pred. No. 0.059;
Matches 61; Conservative 32; Mismatches 93; Indels 64; Gaps 15;

QY 36 TIGGGEYQKKYGGNGLDKLKLYPEIFETFKQVSTGVADPSQKITEKSAKYFNGTWILHR 95
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
13 ting-----kky----yfd-----intgaal--tsykingkhfyndgvmql 50
QY 96 GSGYVLA-KDGGQVYNLGTTTKQFLP-----IQLTGEKK--QGNEGFVKG---N 138
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
51 g---vfkpgdgfeypantqnnieqaiyqskftlogkkyyfdnaskavtgrwlin 107
QY 139 DGNYYFYDLAGNMVKNFFIEDSVGNMYFFDQDGKWE-----NKHFDVDS----YG 186
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
108 nekyfyfn--pnnaiaavglqvldnknkyfnpdtaiskgwqvtngsryyfdtdtaiafng 165

```

QY	187	EK---	GTYFLKNGVSPRGLVDTGNTGYFF---DNYGKWNRQTINAGAMITYTLDENCK	240
Dd	166	yktidgkhfyfdsdcvkigvftstsgsfyfapantynnniegqalvyqskflt1--ngk	223	
QY	241	LIRASYNSDA	250	
Dd	224	kyyfdnnska	233	
RESULT	24			
AAB30543				
ID	AAB30543	standard; Protein; 866 AA.		
AC	AAB30543;			
XX				
DT	06-MAR-2001	(first entry)		
XX				
DE		A recombinant protein of toxin A repeating units (rARU).		
XX				
KW	Toxin A repeating unit; rARU; toxin B repeating unit; rBRU; vaccine;			
KW	passive immune therapy; diarrhoea; colitis.			
OS	Clostridium difficile.			
XX				
FH	Key	Location/Qualifiers		
FT	Misc-difference 794	/note= "c1n encoded by GGTCAA"		
FT	Misc-difference 821..822	/note= "these residues encoded by GGA"		
FT	Misc-difference 836..866	/note= "the nucleotides encoding these residues are not given"		
PN	RQ200061762-A1.			
PD	19-OCT-2000.			
PF	10-APR-2000; 2000WO-US09525.			
PR	09-APR-1999; 99US-0128686.			
PR	01-MAR-2000; 2000US-0186201.			
PR	20-MAR-2000; 2000US-0190111.			
PA	{TECH-} TECHLABS INC.			
PI	Wilkins TD, Lysterly DM, Moncrief JS, Zheng L, Phelps C;			
DR	WPI; 2000-679491/66.			
DR	N-PSDH; AAC62061.			
XX				
PT	Immunogenic composition for use as a vaccine against Clostridium			
PT	difficile, comprises recombinant protein component consisting of at			
PT	least one repeating unit of toxin A or toxin B or both, of Clostridium			
XX	difficile -			
ES	Example 1; Fig 3; 39pp; English.			
CC	The present sequence represents a Clostridium difficile recombinant			
CC	protein comprising toxin A repeating units (rARU). The protein is used			
CC	to prepare recombinant protein components. These recombinant protein			
CC	components comprise at least one repeating unit of Clostridium diffi-			
CC	cile toxin A (rARU), one repeating unit of Clostridium difficile toxin B			
CC	(rBRU), or both. The recombinant protein is used as a vaccine, and is			
CC	useful for conferring a protective response in a mammalian host. It is			
CC	also useful for producing antibodies for passive immune therapy against			
CC	a strain of Clostridium difficile. The vaccines, formulated for use in			
CC	humans and animals, is useful for treating Clostridium difficile			
CC	mediated diarrhoea and colitis.			
SQ	Sequence	866 AA;		

Query Match	8.2%;	Score 117;	DB 21;	Length 866;
Best Local Similarity	24.4%;	Pred. NO. 0.059;		
Matches	61;	Conservative	32;	Mismatches 93;
				Indels 64;
				Gaps
QY	36	TIGGGEYQKRYGGAFLDKLQKLYPEIFTKKQVSTGVAIDPSSOKITWSAKYFNGTNILHR	95	
DB	13	ting-----kky-yfd-----intgaal-tsykingkhfyfndgvmql	50	
QY	96	SGSVLKA-DCGQYYMLGTTTKQFLP-----IQLTGEKK--QGNEGPKVG---N	138	
DB	51	g----vfkpgdgfyfapantqnunieggaivyskfltnkgkkyfdnnskavtgvrriln	107	
QY	139	DGNYFYFDLAGNMVKNFTFIEDSGNMVYFDDGKWE-----NKHFDVDVDS---YG	186	
DB	108	nekyyfn--pnnaiaavglvndnkyyfnpdtaiskwqwtvngryyfdtdafafng	165	
QY	187	EK---GNYFFLKNGVSFRGLGVTDNGTYVF---DNYGKXVRNQTINAGAMITYLDENGK	240	
DB	166	yktidgkhfyfdscvkvigfstngsfyfapantynnnieggaivyskfltn--ngk	223	
QY	241	LIRASYNDSA	250	
DB	224	kyyfdnnksa	233	
RESULT	25			
AAB30546				
ID	AAB30546	standard; Protein: 866	AA.	
XX				
AC	AAB30546;			
XX				
DT	06-MAR-2001	(first entry)		
XX				
DE		A recombinant protein of toxin A repeating units (rARU).		
XX				
KW	Toxin A; repeating unit; rARU; immunogenic composition; polysaccharide			
KW	pathogenic microorganism; nosocomial pathogenic microorganism; vaccine			
KW	immune response.			
XX				
OS	Clostridium difficile.			
XX				
PH	Key	Location/Qualifiers		
FT	Misc-difference 794	/note= "Gln encoded by GGTCAA"		
FT	Misc-difference 821..822	/note= "these residues encoded by GGA"		
FT	Misc-difference 836..866	/note= "the nucleotides encoding these residues are not given"		
XX				
PN	WO200061761-A2.			
XX				
PD	19-OCT-2000.			
XX				
PF	10-APR-2000; 2000WO-US09523.			
XX				
PR	09-APR-1999; 99US-0128686.			
PR	01-MAR-2000; 2000US-0186201.			
XX				
PA	(TECH-) TECHLAB INC.			
PA	(USSH) US DEPT HEALTH & HUMAN SERVICES.			
XX				
PI	Wilkins TD, Lylely DM, Moncrief JS, Pavlakova D, Scheerson R;			
PI	Robbins JB;			
XX				
DR	WPI; 2000-679490/66.			
XX				
PT	Immunogenic compositions useful as vaccines comprise a recombinant			
PT	protein of toxin A or B of Clostridium difficile conjugated to a			
PT	polysaccharide of a microbial pathogen			
XX				
XX	Example 1; Fig 3; 45pp; English.			
XX				

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 12, 2002, 08:02:38 ; Search time 13.2 Seconds
(without alignments)
494.063 Million cell updates/sec

Title: US-09-995-749A-2_COPY_1515_1781

Perfect score:

Sequence: 1 IYNLPGKEVATVTRVDDRG.....SDAEYPTSTDVGKMLDQNK 267

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%

100% Processing: Minimum Match 0%
Maximum Match 100%

Listing first 100 summaries

Database : Issued_Patents_AA:*

```

1: /cqn2_6/ptodata/2/1aa/5A COMB.per.*

```

```
1: /cqn2_v/ptodata/z/1aa/3A_COMB.pep.*
2: /cqn2_6/ptodata/2/1aa/5B_COMB.pep.*
```

```

2: /cgn2_6/ptodata/2/iaa/52_50mb.pcp:
3: /cgn2_6/ptodata/2/iaa/56A_COMB.pcp:

```

4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*

5: /cgn2_6/ptodata/2/laa/PCTUS_COMB.pe

6: /cgn2_6/ptodata/2/1aa/backfiles1.pe

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	510.5	35.7	1430	3	US-09-008-172-2	Sequence 2, Appli
2	510.5	35.7	1430	4	US-09-210-361-6	Sequence 6, Appli
3	458.5	32.1	1475	3	US-09-007-999-2	Sequence 2, Appli
4	458.5	32.1	1475	4	US-09-210-361-2	Sequence 2, Appli
5	445.5	31.2	1375	4	US-09-210-361-4	Sequence 4, Appli
6	418	29.2	1577	2	US-08-793-824-2	Sequence 2, Appli
7	134.5	9.4	591	4	US-08-961-083-74	Sequence 74, Appli
8	128	9.0	569	4	US-08-961-083-154	Sequence 154, Appli
9	125	8.7	2710	1	US-08-480-604A-6	Sequence 6, Appli
10	125	8.7	2710	2	US-08-405-496A-6	Sequence 6, Appli
11	125	8.7	2710	4	US-08-915-136-6	Sequence 6, Appli
12	119.5	8.4	239	4	US-09-485-885-12	Sequence 12, Appli
13	119.5	8.4	292	4	US-09-485-885-10	Sequence 10, Appli
14	119.5	8.4	390	4	US-09-485-885-14	Sequence 14, Appli
15	115.5	8.1	511	1	US-08-480-604A-20	Sequence 20, Appli
16	115.5	8.1	511	2	US-08-405-496A-20	Sequence 20, Appli
17	115.5	8.1	511	4	US-08-915-136-20	Sequence 20, Appli
18	115.5	8.1	608	1	US-08-480-604A-21	Sequence 21, Appli
19	115.5	8.1	608	2	US-08-405-496A-21	Sequence 21, Appli
20	115.5	8.1	608	4	US-08-915-136-21	Sequence 21, Appli
21	115.5	8.1	609	1	US-08-480-604A-30	Sequence 30, Appli
22	115.5	8.1	609	4	US-08-915-136-30	Sequence 30, Appli
23	115.5	8.1	2366	1	US-08-480-604A-10	Sequence 10, Appli
24	115.5	8.1	2366	2	US-08-405-496A-10	Sequence 10, Appli
25	115.5	8.1	2366	4	US-08-915-136-10	Sequence 10, Appli
26	114.5	8.0	631	4	US-08-847-065-25	Sequence 25, Appli
27	113	7.9	220	4	US-09-308-022-9	Sequence 9, Appli

ALIGNMENTS

RESULT 1

US-09-008-172-2
; Sequence 2, Application US/09008172
; Patent No. 6127602
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starch and
; FILE REFERENCE: 0358D
; CURRENT APPLICATION NUMBER: US/09/008,172
; CURRENT FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1430
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-008-172-2

Query Match 35.7%; Score 510.5; DB 3; Length 1430;
Best Local Similarity 43.6%; Pred. No. 1.5e-38;
Matches 120; Conservative 33; Mismatches 79; Indels 43; Gaps 12;
QY 1 IYNLPGEVATVTRVDDRGVNWKDAIINNLYVNTTGGG-EYOKYGGAFDLKLOKLYP 59
Db IYLPGEVATVTRVDDRGVNWKDAIINNLYVNTTGGG-EYOKYGGAFDLKLOKLYP 59
QY 60 EIFTKKQVSTGVAIDPSOKITENSAYFNGTNIHNRSGVYVLKADGG-QYINL-GTTTQ 117
Db IYLPGEVATVTRVDDRGVNWKDAIINNLYVNTTGGG-EYOKYGGAFDLKLOKLYP 59
QY 1026 SIFNRTQISNGKIDPSEKITAANKYFNGTNIHNRSGVYVLKADGG-QYINL-GTTTQ 117
Db IYLPGEVATVTRVDDRGVNWKDAIINNLYVNTTGGG-EYOKYGGAFDLKLOKLYP 59
QY 118 FLPIQLTGEKKQNGEYFNGTNIHNRSGVYVLKADGG-QYINL-GTTTQ 117
Db IYLPGEVATVTRVDDRGVNWKDAIINNLYVNTTGGG-EYOKYGGAFDLKLOKLYP 59
QY 1084 YLPKQMT--NKEASTGFV--NDGNGMTFYSTGYOAKNSFVQDAGNMYFDDNNGHMYG 1139
Db IYLPGEVATVTRVDDRGVNWKDAIINNLYVNTTGGG-EYOKYGGAFDLKLOKLYP 59
QY 177 KHFVDVDSYGEKGYFFELKNGVSFRGGLVQT-----DNGTYFYFDNYGK-- 219
Db IYLPGEVATVTRVDDRGVNWKDAIINNLYVNTTGGG-EYOKYGGAFDLKLOKLYP 59
QY 1140 LQQLN---GE--VOYFSLNGVQLRESFLENADGSKNYPGHLGNRYSNGYFSDNDSKWR 1193
Db IYLPGEVATVTRVDDRGVNWKDAIINNLYVNTTGGG-EYOKYGGAFDLKLOKLYP 59
QY 220 -----MVRNQTINAGAMYITLDENGLIKASY 246
Db IYLPGEVATVTRVDDRGVNWKDAIINNLYVNTTGGG-EYOKYGGAFDLKLOKLYP 59
QY 1194 YFDASGVMAVGLKTINGNTQYF--DQDGYQVKGAW 1226
Db IYLPGEVATVTRVDDRGVNWKDAIINNLYVNTTGGG-EYOKYGGAFDLKLOKLYP 59

RESULT 2

US-09-210-361-6
; Sequence 6, Application US/09210361
; Patent No. 6284479
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starches and
; FILE REFERENCE: 0357CR
; CURRENT APPLICATION NUMBER: US/09/210,361
; CURRENT FILING DATE: 1998-12-11
; EARLIER APPLICATION NUMBER: 09/007,999
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/009,620
; EARLIER FILING DATE: 1998-01-20
; EARLIER APPLICATION NUMBER: 08/485,243
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/008,172
; EARLIER FILING DATE: 1998-01-16

; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 1430
; TYPE: PRT
; ORGANISM: streptococcus mutans
US-09-210-361-6

Query Match 35.7%; Score 510.5; DB 4; Length 1430;
Best Local Similarity 43.6%; Pred. No. 1.5e-38;
Matches 120; Conservative 33; Mismatches 79; Indels 43; Gaps 12;
QY 1 IYNLPGEVATVTRVDDRGVNWKDAIINNLYVNTTGGG-EYOKYGGAFDLKLOKLYP 59
Db IYLPGEVATVTRVDDRGVNWKDAIINNLYVNTTGGG-EYOKYGGAFDLKLOKLYP 59
QY 966 IYNLPGEVATVTRVDDRGVNWKDAIINNLYVNTTGGG-EYOKYGGAFDLKLOKLYP 59
Db IYLPGEVATVTRVDDRGVNWKDAIINNLYVNTTGGG-EYOKYGGAFDLKLOKLYP 59
QY 60 EIFTKKQVSTGVAIDPSOKITENSAYFNGTNIHNRSGVYVLKADGG-QYINL-GTTTQ 117
Db IYLPGEVATVTRVDDRGVNWKDAIINNLYVNTTGGG-EYOKYGGAFDLKLOKLYP 59
QY 1026 SIFNRTQISNGKIDPSEKITAANKYFNGTNIHNRSGVYVLKADGG-QYINL-GTTTQ 117
Db IYLPGEVATVTRVDDRGVNWKDAIINNLYVNTTGGG-EYOKYGGAFDLKLOKLYP 59
QY 118 FLPIQLTGEKKQNGEYFNGTNIHNRSGVYVLKADGG-QYINL-GTTTQ 117
Db IYLPGEVATVTRVDDRGVNWKDAIINNLYVNTTGGG-EYOKYGGAFDLKLOKLYP 59
QY 1084 YLPKQMT--NKEASTGFV--NDGNGMTFYSTGYOAKNSFVQDAGNMYFDDNNGHMYG 1139
Db IYLPGEVATVTRVDDRGVNWKDAIINNLYVNTTGGG-EYOKYGGAFDLKLOKLYP 59
QY 177 KHFVDVDSYGEKGYFFELKNGVSFRGGLVQT-----DNGTYFYFDNYGK-- 219
Db IYLPGEVATVTRVDDRGVNWKDAIINNLYVNTTGGG-EYOKYGGAFDLKLOKLYP 59
QY 1140 LQQLN---GE--VOYFSLNGVQLRESFLENADGSKNYPGHLGNRYSNGYFSDNDSKWR 1193
Db IYLPGEVATVTRVDDRGVNWKDAIINNLYVNTTGGG-EYOKYGGAFDLKLOKLYP 59
QY 220 -----MVRNQTINAGAMYITLDENGLIKASY 246
Db IYLPGEVATVTRVDDRGVNWKDAIINNLYVNTTGGG-EYOKYGGAFDLKLOKLYP 59
QY 1194 YFDASGVMAVGLKTINGNTQYF--DQDGYQVKGAW 1226
Db IYLPGEVATVTRVDDRGVNWKDAIINNLYVNTTGGG-EYOKYGGAFDLKLOKLYP 59

RESULT 3

US-09-007-999-2
; Sequence 2, Application US/09007999
; Patent No. 6087559
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starch and
; FILE REFERENCE: 0356D
; CURRENT APPLICATION NUMBER: US/09/007,999
; CURRENT FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-007-999-2

Query Match 32.1%; Score 458.5; DB 3; Length 1475;
Best Local Similarity 38.9%; Pred. No. 9.7e-34;
Matches 111; Conservative 39; Mismatches 94; Indels 41; Gaps 12;
QY 1 IYNLPGEVATVTRVDDRGVNWKDAIINNLYVNTTGGG-EYOKYGGAFDLKLOKLYP 59
Db IYLPGEVATVTRVDDRGVNWKDAIINNLYVNTTGGG-EYOKYGGAFDLKLOKLYP 59
QY 934 MIAPPEKEVATVTRVDDRGVNWKDAIINNLYVNTTGGG-EYOKYGGAFDLKLOKLYP 59
Db IYLPGEVATVTRVDDRGVNWKDAIINNLYVNTTGGG-EYOKYGGAFDLKLOKLYP 59
QY 60 EIFTKKQVSTGVAIDPSOKITENSAYFNGTNIHNRSGVYVLKADGG-QYINL-GTTTQ 117
Db IYLPGEVATVTRVDDRGVNWKDAIINNLYVNTTGGG-EYOKYGGAFDLKLOKLYP 59
QY 994 ELFARKQISTGVPMDSVKIKQNSAKYFNGTNIHNRSGVYVLKADGG-QYINL-GTTTQ 117
Db IYLPGEVATVTRVDDRGVNWKDAIINNLYVNTTGGG-EYOKYGGAFDLKLOKLYP 59
QY 118 FLPIQLTGEKKQNGEYFNGTNIHNRSGVYVLKADGG-QYINL-GTTTQ 117
Db IYLPGEVATVTRVDDRGVNWKDAIINNLYVNTTGGG-EYOKYGGAFDLKLOKLYP 59
QY 1054 FLPKTLNQDSQ--VGF-SYDCKG--YVYVSTGYOAKNTFISEG-DKWWYFDNNGYHWGTA 1109
Db IYLPGEVATVTRVDDRGVNWKDAIINNLYVNTTGGG-EYOKYGGAFDLKLOKLYP 59

```
QY 178 HFVDVDSGKGTGYFFLKNVSGRGLVQTDNGTY-YFDNYGKVRN-----QTI 226
Db 1110 QSIN-----GVNIYFLNSQLRDAILKNEGTYAYYGNDRYENGYYQFMGVRHFE 1163
QY 227 NAGAM-----IYTLDENGKLKASYNDAEYPTSTDVGKM 261
Db 1164 NNGEMSVGLTVIDGQVYFDEMG-----YQAKGKFVTTAD-GKI 1201

RESULT 4
US-09-210-361-2
; Sequence 2, Application US/09210361
; Patent No. 6284479
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starches and
; FILE OF INVENTION: Latexes in Paper Manufacture
; FILE REFERENCE: 0357CR
; CURRENT APPLICATION NUMBER: US/09/210,361
; EARLIER APPLICATION NUMBER: 09/007,999
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/009,620
; EARLIER FILING DATE: 1998-01-20
; EARLIER APPLICATION NUMBER: 08/485,243
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/008,172
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
; US-09-210-361-2

Query Match 32.1%; Score 458.5; DB 4; Length 1475;
Best Local Similarity 38.9%; Pred. No. 9.7e-34;
Matches 111; Conservative 39; Mismatches 94; Indels 41; Gaps 12;

QY 1 IYNLPKEVATVTRVDRGNVWDAIINNLYVN-TIGGGEYKKGAFDLKQLKP 59
Db 934 MYALPEKEVATVTRVDRGNVWDAIINNLYVN-TIGGGEYKKGAFDLKQLKP 59
QY 60 EIFTKQVSTGVAIDPSQKITERSAKYFNGTNIHRGSGVVLKADG-QYVNLGTTK-Q 117
Db 994 ELPAKQISTGVPMDFSVKIKQWSAKYFNGTNIHRGSGVVLKADG-QYVNLGTTK-Q 1053
QY 118 FLPTQLTGKQKQNEGVKNGDGNFYFDLAGNKNVNTFIEDSGVGNVYFFDQGGKVENK 177
Db 1054 FLPTLLNQDSQ--VGSYDQKQ-YVYVSTSGYQAKNTFISEG-DKWFYFDNNGYWTGA 1109
QY 178 HFVDVDSGKGTGYFFLKNVSGRGLVQTDNGTY-YFDNYGKVRN-----QTI 226
Db 1110 QSIN-----GVNIYFLNSQLRDAILKNEGTYAYYGNDRYENGYYQFMGVRHFE 1163
QY 227 NAGAM-----IYTLDENGKLKASYNDAEYPTSTDVGKM 261
Db 1164 NNGEMSVGLTVIDGQVYFDEMG-----YQAKGKFVTTAD-GKI 1201

RESULT 5
US-09-210-361-4
; Sequence 4, Application US/09210361
; Patent No. 6284479
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
```

```
; TITLE OF INVENTION: Substitutes for Modified Starches and
; FILE OF INVENTION: Latexes in Paper Manufacture
; FILE REFERENCE: 0357CR
; CURRENT APPLICATION NUMBER: US/09/210,361
; EARLIER APPLICATION NUMBER: 09/007,999
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/009,620
; EARLIER FILING DATE: 1998-01-20
; EARLIER APPLICATION NUMBER: 08/485,243
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/008,172
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 1375
; TYPE: PRT
; ORGANISM: streptococcus mutans
; US-09-210-361-4

Query Match 31.2%; Score 445.5; DB 4; Length 1375;
Best Local Similarity 39.9%; Pred. No. 1.4e-32;
Matches 115; Conservative 32; Mismatches 96; Indels 45; Gaps 13;

QY 1 IYNLPKEVATVTRVDRGNVWDAIINNLYVN-TIGGGEYKKGAFDLKQLKP 59
Db 961 MYALPEKEVATVTRVDRGNVWDAIINNLYVN-TIGGGEYKKGAFDLKQLKP 59
QY 60 EIFTKQVSTGVAIDPSQKITERSAKYFNGTNIHRGSGVVLKADG-QYVNLGTTK-Q 119
Db 1021 ELPAKQISTGVPMDFSVKIKQWSAKYFNGTNIHRGSGVVLKADG-QYVNLGTTK-Q 1080
QY 120 PIQLTGKQKQNEGVKNGDGNFYFDLAGNKNVNTFIEDSVG-NWYFFDQGGKWE 175
Db 1081 PKSLV-NPNHGTSSVTGLVFDGKGYVYSTSGNQAKNAFI--SLGNWYFDNNGYWT 1137
QY 176 NKHFVDVDSGKGTGYFFLKNVSGRGLVQTDNGTY-YFDNYGKVRN-----QTI 213
Db 1138 GASIN-----GANYFELNSQLRDAILKNEGTYAYYGNDRYENGYYQFMGVRHFE 1191
QY 214 FUNDYGNVNRQNTNAGAMVYTLDENGKLKASYNDAEYPTSTDVGKM 261
Db 1192 FQN-GIMAVGLTVRHSVAVQY-FDASG-----FOAKGQFVTTAD-GKI 1230

RESULT 6
US-08-793-824-2
; Sequence 2, Application US/08793824
; Patent No. 5961838
; GENERAL INFORMATION:
; APPLICANT: Simpson, Christine Lynn
; APPLICANT: Giffard, Philip Morrison
; APPLICANT: Jacques, Nicholas Anthony
; TITLE OF INVENTION: Genetic Manipulation of Plants to
; TITLE OF INVENTION: Increase Stored Carbohydrates
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESS: Griffith Hack & Co
; CITY: Level 8, 168 Walker Street
; CITY: No. 5981838th Sydney
; STATE: New South Wales
; COUNTRY: Australia
; ZIP: 2060
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
```


;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/793,824
;; FILING DATE:
;; CLASSIFICATION: 800
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: AU PM7643
;; FILING DATE: 24-AUG-1994
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 61 2 9957 5944
;; TELEFAX: 61 2 957 6288
;; TELEX: 36547
;; INFORMATION FOR SEQ ID NO: 2:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1577 amino acids
;; TYPE: amino acid
;; STRANDEDNESS:
;; TOPOLOGY: not relevant
;; MOLECULE TYPE: protein
;; ORIGINAL SOURCE:
;; ORGANISM: Streptococcus salivarius
US-08-793-824-2

Query Match 29.2%; Score 418; DB 2; Length 1577;
Best Local Similarity 39.0%; Pred. No. 5.8e-30;
Matches 105; Conservative 38; Mismatches 82; Indels 44; Gaps 11;
QY 1 IYNLPCKEVATVTRVDRGNVWKAIDAIINNLLYVNT-IGGGEYQKYGGAFLDKLQKLYP 59
Db 1042 IYNLPCKEVATVTRVDRGNVWKAIDAIINNLLYVNT-IGGGEYQKYGGAFLDKLQKLYP 1101
QY 60 EIPTKQVSTGVAIDQSOKITWSAKYFNGTNILHRGSGYVLKADG-QQYNL--GTTTK 116
Db 1102 AIFERVQISNGRKLATNEKITOMSAKYFNGTNILHRGSGYVLKADG-QQYNL--GTTTK 1159
QY 117 QFLPIQLTCEKKGNGGVK-GNDGNYFYDLAGNMVKNFTIEDSVGNMYFFDQDGKAYE 175
Db 1160 -FLPKQMT--EITGSGFRVGDVQY--LSTGCLAKNTFFIQVGNQHYEYDKNGNWT 1213
QY 176 NKHFVDVDSYGEKGTFFELKNGVSFSGGL-----VQTDNGTY----- 212
Db 1214 GEQVD-----GKK--YFFLDNGLQLRHLVLRQSDCHVYVDPKGVQAFNGFYDFAGPROD 1267
QY 213 --YFDNYGKVMNQTNAGAMIYTLDENG 239
Db 1268 VRYFDNGQMYRGLHDMYGTTFYDEKTC 1296

RESULT 7
US-08-961-083-74
; Sequence 74, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigen and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,083
; FILING DATE:
; CLASSIFICATION: 435

;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER:
;; FILING DATE:
;; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 74:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 591 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-961-083-74

Query Match 9.4%; Score 134.5; DB 4; Length 591;
Best Local Similarity 20.6%; Pred. No. 0.00021;
Matches 64; Conservative 48; Mismatches 112; Indels 87; Gaps 14;
QY 2 YNLPCKEVATVTRVDRGNVW---KDAIINNLLYVNTIGGGE-----YQKKGGA 49
Db 54 YLKSGGYMAKSEWEDKGFYLLDQDGRKMRNMTVGTGATGAKVIEQWVYDSQYDAM 113
QY 50 FLDKLQKLYPEITFKKQVSTGVAIDPSOKITWSAKYFNGTNILHRGSGYVLKADGQY 109
Db 114 FYKADGQHAEE-----NEW--LQIKGNDYFKSGYLLTS---QWI 149
QY 110 NLCTTKQFLPTQ--LTCBKKGNGFVKGN-----DGNYYFDLAGNMVKNFT 156
Db 150 NQAYYNASAKYVQDQWLFQYQSWFYIKENGNYADKEWIFENGHYIYKSGGYMAANE 209
QY 157 IDSNGNWEFFDQDGKVMENKHFYVDVDSVG-----EKGYEFLKNGVS 199
Db 210 INDK-ESWEYLFKDGKMAKEKWEVYDSHQAYYFKSGGYMTANEWIDKESWYFLKSD-- 266
QY 200 FRGLVQTD-----NCTYYFDNYGKVMNQTNAGAMIYTLDENGKLIKA--SYNSDAE 251
Db 267 --GKIAKEWYVDSHQAYYFKSGGYMAKNETVDC----YQLGSDGKWLGGKTTNENAA 320
QY 252 Y---PTSDV 258
Db 321 YQVVPVTANV 331

RESULT 8
US-08-961-083-154
; Sequence 154, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigen and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,083
; FILING DATE:
; CLASSIFICATION: 435

```
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/422,711
; FILING DATE: 14-APR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 154:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 569 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-961-083-154

Query Match          9.0%; Score 128; DB 4; Length 569;
Best Local Similarity 22.8%; Pred. No. 0.00081;
Matches 59; Conservative 41; Mismatches 109; Indels 50; Gaps 13;

QY 37 IGGGEYQKYGGAFLDK--LQKLYPEIFTKKOVSTGVAIDPSQKITEW-----SAKYENG 89
Db 69 IRGDIYFKSGGYLLTSQINQAVNNSGAK-VQGWLFDP--KQYSWFYIKENGNYADK 125
QY 90 TNLHRSQGVVLKADGGQYINLTGTTKQ-FLPIQLTGERKKQGNQGVFKNGDGNFYFDILA 148
Db 126 EWIFENGHYILKSGGYMAANEMWIDKESWFLKPDG-KMAEKWYDSDSHSQAWYFKSG 184
QY 149 GRVKKWTFEDSVGNWYFDQDGKAVENKHFVDVDSYG-----EKGTY 191
Db 185 GYMANEWIWDK-ESWFLKSDGKIAEKWYDSDSHSQAWYFKSGGYMTANEMWIDKESW 243
QY 192 FFLKNGVSPRGLVQFD-----NGTYFDNGYKMKVRNOTINAGAMITVLDENGKLKA 244
Db 244 FYLKSD---GKIAEKWYDSDSHSQAWYFKSGGYMAKNETVDG----YQLGSDGKWLGG 295
QY 245 -SYNSDAEY-----PTSDV 258
Db 296 KTTNENAAYYQVVPVTAVV 314

RESULT 9
US-08-480-604A-6
; Sequence 6, Application US/08480604A
; Patent No. 5736139
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALLEY, BRUCE S.
; APPLICANT: PADHYE, NISHA V.
; APPLICANT: FINCA, JOSEPH R.
; APPLICANT: STAFFORD, DOUGLAS C.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; PREVENTION OF C. DIFFICILE DISEASE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,604A
; FILING DATE: 07-JUN-1995
```

```
;
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/422,711
; FILING DATE: 14-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/405,496
; FILING DATE: 16-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-01763
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2710 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-480-604A-6

Query Match          8.7%; Score 125; DB 1; Length 2710;
Best Local Similarity 24.1%; Pred. No. 0.012;
Matches 68; Conservative 37; Mismatches 101; Indels 76; Gaps 17;

QY 16 DRGNVWKDAI-INNNLYVN-----TIGGEYQKYGGAFLDKLQKLYPEIFT 63
Db 1824 DEDSKLVKGLININNSLFYFDPDEFNLVTGWOTING----KKY----YFD----- 1865
QY 64 KQOVSTGVAIDPSQKITEWSAKYENGNTLHRSQGVVLKA-DGGQYINLTGTTKQFLP-- 120
Db 1866 ---INGAAL-TSYKIINGKHFFNNDGVNQIG---VFKGPDGFETAFAPANTQNNIEGQ 1918
QY 121 -----IQLTGEKK--Q3NEGFKVG-----NDGNYFYDLAGNMAVKKTFIEDSVGNWYF 166
Db 1919 AIYQSKFLTLNGKKYFEDNNSKAVTGWRIINNEKYFN--PNNATAAVGLQVIDNNKY 1976
QY 167 FDQDGKAWB-----NKHFDVDS-----YGEK---GYFFLKNGVSPRGLVQDTNGT 211
Db 1977 FNPDTAILSKGWQTVNGSRYFYDITDAIFNGYKTDGKHFFEDSDCVVYKIGVFTSNGF 2036
QY 212 YVF---DNYGKMYRQNTINAGAMITVLDENGKLKASYNDA 250
Db 2037 EYFAPANTNNIEGQAIYQSKFLTL--NGKKYFDNNSKA 2076

RESULT 10
US-08-405-496A-6
; Sequence 6, Application US/08405496A
; Patent No. 5919665
; GENERAL INFORMATION:
; APPLICANT: WILLIAMS, JAMES A.
; TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM
; TITLE OF INVENTION: NEUROTOXIN
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
```


APPLICANT: KINK, JOHN A.
APPLICANT: THALLEY, BRUCE S.

APPLICANT: KINK, JOHN A.
APPLICANT: THALLEY, BRUCE S.

APPLICANT: KINK, JOHN A.
APPLICANT: THALLEY, BRUCE S.

APPLICANT: PADHYE, NISHA V.
APPLICANT: FIRCA, JOSEPH R.
APPLICANT: STAFFORD, DOUGLAS C.
TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
PREVENTION OF C. DIFFICILE DISEASE
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEN & CARROLL, LLP
STREET: 220 MONTGOMERY STREET, SUITE 2200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 94104

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US 08/480,604A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/422,711
FILING DATE: 14-APR-1995

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/405,496
FILING DATE: 16-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329,154
FILING DATE: 25-OCT-1994

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161,907
FILING DATE: 02-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/985,321
FILING DATE: 04-DEC-1992

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/429,791
FILING DATE: 31-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: INGOLIA, DIANE E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPHD-01763

TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 511 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-480-604A-20

Query Match 8.1%; Score 115.5; DB 1: Length 511;
Best Local Similarity 24.5%; Pred. No. 0.0099;
Matches 68; Conservative 39; Mismatches 102; Indels 69; Gaps 15;

QY 27 INNLYVNTIGGGEYQKKYGAFDLKLYPEFTKKQV-STGVAIDPSQKITEWSAK 85
DB 8 VGDDKTYFNPIGNG--AASIGETIID--DRNY--YFNQSGVLQTVGF-----STEDGFK 55
QY 86 YFNGTNIIL-----HRSGYVLKADGGQYVNLGTTTKQF 118
DB 56 YFANTLDENLEGEAIDFTCKLIIDENIYFDNYSRGAWEKELDCEMHYFSPETGKAF 115
QY 119 LPQIGTEKKO-----GNBGFVKGNNGHYFDLAGNMVKNFTFIEDSGVNGVYFFPDQ 171
DB 116 KGLNQIGDYKYFNDSGVQKGFVSLDNKXHYFDGSGVMKVGYTEID---GKHYPFAENG 172

QY 172 KM---VEN-----KHFV--DVDSYGKGTFFLKNKGVSRGGLVQTDNGTYFP-DNYGKM 220
DB 173 EMQIGVFTEDGKFFAHHNEDLNEEG-----EISY-SGILFNKNIYFFDSDFTAV 225
QY 221 VNOQTINAGAMITYLDENCK-LIKASVNSDAEYPTSTD 257
DB 226 VGWKLEDSGSKYFFEDTAETAYIGLSLINDGQYFND 263

RESULT 16

US-08-405-496A-20
; Sequence 20, Application US/08405496A
; Patent No 5919665
; GENERAL INFORMATION:

; APPLICANT: WILLIAMS, JAMES A.
; TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94104

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/405,496A
FILING DATE: 16-MAR-1995
CLASSIFICATION: 424

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329,154
FILING DATE: 25-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161,907
FILING DATE: 02-DEC-1993

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/985,321
FILING DATE: 04-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/429,791
FILING DATE: 31-OCT-1989

ATTORNEY/AGENT INFORMATION:
NAME: INGOLIA, DIANE E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPHD-01308
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 511 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-405-496A-20

Query Match 8.1%; Score 115.5; DB 2: Length 511;
Best Local Similarity 24.5%; Pred. No. 0.0099;
Matches 68; Conservative 39; Mismatches 102; Indels 69; Gaps 15;

QY 27 INNLYVNTIGGGEYQKKYGAFDLKLYPEFTKKQV-STGVAIDPSQKITEWSAK 85
DB 8 VGDDKTYFNPIGNG--AASIGETIID--DRNY--YFNQSGVLQTVGF-----STEDGFK 55
QY 86 YFNGTNIIL-----HRSGYVLKADGGQYVNLGTTTKQF 118

Db 56 YFANTLDENLSEGAIDFTGKLIIDENIYFDNRYKAVENKELDGENHYFSPETGKAF 115
QY 119 LPIQLGGEKKQ-----GNEGFKVKGNDGNYFYDLAGNMVKNFTIEDSVGNMYFFDDQG 171
Db 116 KGLNQIGDYKYFNSDGMQKGFVINDNKHEDDSDGMKVGYTEID--GKHFFAENG 172
QY 172 KM---VEN-----KHFV--DVDSYGEKGYFFELKNGVSFRGGLVQTDNGTYF-DNYGKM 220
Db 173 EMQIGVENTEDGFKYFAHNEDLGNEG-----EISY-SGILNFKNKIYFDDSDFTAV 225
QY 221 VRNQTINAGAMIYTLDENGK-LIKASYNDAEYPTSTD 257
Db 226 VGMKLDGSKYFDEDTARAYIGLSLINDGQYFENDD 263

RESULT 17

US-08-915-136-20
Sequence 20, Application US/08915136
Patent No. 6290960
GENERAL INFORMATION:
APPLICANT: KINK, JOHN A.
APPLICANT: THALLEY, BRUCE S.
APPLICANT: PADHYE, NISHA V.
APPLICANT: FIRCA, JOSEPH R.
APPLICANT: STAFFORD, DOUGLAS C.
TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
PREVENTION OF C. DIFFICILE DISEASE
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEN & CARROLL, LLP
STREET: 220 MONTGOMERY STREET, SUITE 2200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 94104

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/915,136
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/480,604
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/405,496
FILING DATE: 16-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329,154
FILING DATE: 25-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161,907
FILING DATE: 02-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/985,321
FILING DATE: 04-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/429,791
FILING DATE: 31-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: INGOLIA, DIANE E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPHD-01763
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 511 amino acids
TYPE: amino acid

STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-915-136-20
Query Match 8.1%; Score 115.5; DB 4; Length 511;
Best Local Similarity 24.5%; Pred. NO. 0.0099;
Matches 68; Conservative 39; Mismatches 102; Indels 69; Gaps 15;
QY 27 INNLYVNTIGGGRYQKKYGAFLDKLQKLYPEITKKQV-STGVAIDPSQKITWSAK 85
Db 8 VGDDKYFENPING--AASIGETIID--DKNY--YFNQSCVLQTVF-----STEDGFK 55
QY 86 YFNGTNIL-----HRSGYVLKADGGQYNNLGTPTKQF 118
Db 56 YFAPANTLDENLSEGAIDFTGKLIIDENIYFDNRYKAVENKELDGENHYFSPETGKAF 115
QY 119 LPIQLGGEKKQ-----GNEGFKVKGNDGNYFYDLAGNMVKNFTIEDSVGNMYFFDDQG 171
Db 116 KGLNQIGDYKYFNSDGMQKGFVINDNKHEDDSDGMKVGYTEID--GKHFFAENG 172
QY 172 KM---VEN-----KHFV--DVDSYGEKGYFFELKNGVSFRGGLVQTDNGTYF-DNYGKM 220
Db 173 EMQIGVENTEDGFKYFAHNEDLGNEG-----EISY-SGILNFKNKIYFDDSDFTAV 225
QY 221 VRNQTINAGAMIYTLDENGK-LIKASYNDAEYPTSTD 257
Db 226 VGMKLDGSKYFDEDTARAYIGLSLINDGQYFENDD 263

RESULT 18

US-08-480-604A-21
Sequence 21, Application US/08480604A
Patent No. 5736139
GENERAL INFORMATION:
APPLICANT: KINK, JOHN A.
APPLICANT: THALLEY, BRUCE S.
APPLICANT: PADHYE, NISHA V.
APPLICANT: FIRCA, JOSEPH R.
APPLICANT: STAFFORD, DOUGLAS C.
TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
PREVENTION OF C. DIFFICILE DISEASE
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEN & CARROLL, LLP
STREET: 220 MONTGOMERY STREET, SUITE 2200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/480,604A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/422,711
FILING DATE: 14-APR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/405,496
FILING DATE: 16-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329,154
FILING DATE: 25-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161,907
FILING DATE: 02-DEC-1993
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/985,321
FILING DATE: 04-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/429,791
FILING DATE: 31-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: INGOLIA, DIANE E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPD-01763
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 608 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-480-604A-21

Query Match 8.1%; Score 115.5; DB 1; Length 608;
Best Local Similarity 24.5%; Pred. No. 0.013;
Matches 68; Conservative 39; Mismatches 102; Indels 69; Gaps 15;

QY 27 INNLYVNTIGGEYQKYGAFDLKQLKLYPEIFTKQY-STGVAIDPSQKITEWSAK 85
DB 105 VGDDKYFYFNPINGG--AASIGETIID--DKNY--YFNQSGVLTQGVF-----STEDGFK 152
QY 86 YFNGTNI-----HRGSGVVLKADGGYYNLTGTTKQF 118
DB 153 YFAPANTLDENLEGEAIDFTGKLIIDENIYFDNRYGAVENKELDGEMHYFSPETGKAF 212
QY 119 LPIQLTGEKKQ-----GNBGFVKGNDGNYFYDLAGNMVKNFTFIEDSGNMWYFFDQDG 171
DB 213 KGLNQIGDYKYFYFNSDGVMOGKFVSIINDNKHVFDSDGVKMGVGYTEID---GKHIFYFAENG 269
QY 172 KM---VEN-----KHVV--DVDSYGEKGYFFFLKNGVSFRGGLVQTDNGTYYP-DNYGKM 220
DB 270 EMQIGVFTEDGKIFYFAHNEDLGNEEG-----EISY-SGILNFNKNIYYFDSDSTAV 322
QY 221 VRNQTINAGMIYTLDENGK-LIKASVNSDAEYPTSTD 257
DB 323 VGKQLEDGSGKYFYFDETAAYIGLSLINDGQYYFND 360

RESULT 19
US-08-405-496A-21
Sequence 21, Application US/08405496A
Patent No. 5919665
GENERAL INFORMATION:
APPLICANT: WILLIAMS, JAMES A.
TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM
TITLE OF INVENTION: NEUTROXIN
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEN & CARROLL, LLP
STREET: 220 MONTGOMERY STREET, SUITE 2200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/405,496A
FILING DATE: 16-MAR-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/329,154
FILING DATE: 25-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161,907
FILING DATE: 02-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/985,321
FILING DATE: 04-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/429,791
FILING DATE: 31-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: INGOLIA, DIANE E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPD-01308
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 608 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-405-496A-21

Query Match 8.1%; Score 115.5; DB 2; Length 608;
Best Local Similarity 24.5%; Pred. No. 0.013;
Matches 68; Conservative 39; Mismatches 102; Indels 69; Gaps 15;
QY 27 INNLYVNTIGGEYQKYGAFDLKQLKLYPEIFTKQY-STGVAIDPSQKITEWSAK 85
DB 105 VGDDKYFYFNPINGG--AASIGETIID--DKNY--YFNQSGVLTQGVF-----STEDGFK 152
QY 86 YFNGTNI-----HRGSGVVLKADGGYYNLTGTTKQF 118
DB 153 YFAPANTLDENLEGEAIDFTGKLIIDENIYFDNRYGAVENKELDGEMHYFSPETGKAF 212
QY 119 LPIQLTGEKKQ-----GNBGFVKGNDGNYFYDLAGNMVKNFTFIEDSGNMWYFFDQDG 171
DB 213 KGLNQIGDYKYFYFNSDGVMOGKFVSIINDNKHVFDSDGVKMGVGYTEID---GKHIFYFAENG 269
QY 172 KM---VEN-----KHVV--DVDSYGEKGYFFFLKNGVSFRGGLVQTDNGTYYP-DNYGKM 220
DB 270 EMQIGVFTEDGKIFYFAHNEDLGNEEG-----EISY-SGILNFNKNIYYFDSDSTAV 322
QY 221 VRNQTINAGMIYTLDENGK-LIKASVNSDAEYPTSTD 257
DB 323 VGKQLEDGSGKYFYFDETAAYIGLSLINDGQYYFND 360

RESULT 20
US-08-915-136-21
Sequence 21, Application US/08915136
Patent No. 6290960
GENERAL INFORMATION:
APPLICANT: KINK, JOHN A.
APPLICANT: THALLEY, BRUCE S.
APPLICANT: PADHYE, NISHA V.
APPLICANT: FIRCA, JOSEPH R.
APPLICANT: STAFFORD, DOUGLAS C.
TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEN & CARROLL, LLP
STREET: 220 MONTGOMERY STREET, SUITE 2200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 94104

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/915,136
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/480,604
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/405,496
FILING DATE: 16-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329,154
FILING DATE: 25-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161,907
FILING DATE: 02-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/985,321
FILING DATE: 04-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/429,791
FILING DATE: 31-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: INGOLIA, DIANE E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPND-01763
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 608 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-915-136-21

Query Match 8.1%; Score 115.5; DB 4; Length 608;
Best Local Similarity 24.5%; Pred. No. 0.013;
Matches 68; Conservative 39; Mismatches 102; Indels 69; Gaps 15;
QY 27 INNLYVNTIGGGEYQKYGAFDLKQKLYPEITFKQV-STGVADIPSQKITEWSAK 85
DB 105 VGDDKYFNPINGG--AASIGETIID--DKNY--YFQSGVLQTVF-----STEDGFK 152
QY 86 YFNGTNIL-----HRGSGYVLKADGGQYVNLGTTTKQF 118
DB 153 YFAPANTLDNLEGEAIDFTGKLIIDENIYFDDNYRGAVENKELGEMHYFSPETGKAF 212
QY 119 LPIQLTGEKKQ-----GNEGFKVGHGNYFYDLAGNMVKNFTIEDSYGNWYFFDQDG 171
DB 213 KGLNQIGDYKYFYFNSDGMQKGFVSINDKNHYFDDSGVMKVGYTEID--GKHFEYFAENG 269
QY 172 KM---VEN-----KHVY--DVDSYGEKGTFFLKNVGSFRGGLVQDTNGTYIF-DNYGKM 220
DB 270 EMQIGVFTNTGDKFYFAHNEEDLGNESG-----EELSY-SGLNFNKNKIYFDDSTAV 322
QY 221 VRNQTINAGMIYTLDENGK-LIKASVNSDAEYPTSTD 257
DB 323 VGMKLEDSGSKYYFDEDTAEAYIGLSLINDGQYFYND 360

RESULT 21
US-08-480-604A-30
Sequence 30, Application US/08480604A
Patent No. 5736139

GENERAL INFORMATION:
APPLICANT: KINK, JOHN A.
APPLICANT: THALLEY, BRUCE S.
APPLICANT: PADHYE, NISHA V.
APPLICANT: FIRCA, JOSEPH R.
APPLICANT: STAFFORD, DOUGLAS C.
TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEN & CARROLL, LLP
STREET: 220 MONTGOMERY STREET, SUITE 2200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/480,604A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/422,711
FILING DATE: 14-APR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/405,496
FILING DATE: 16-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329,154
FILING DATE: 25-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161,907
FILING DATE: 02-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/985,321
FILING DATE: 04-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/429,791
FILING DATE: 31-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: INGOLIA, DIANE E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPND-01763
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 30:
SEQUENCE CHARACTERISTICS:
LENGTH: 609 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-480-604A-30

Query Match 8.1%; Score 115.5; DB 1; Length 609;
Best Local Similarity 24.5%; Pred. No. 0.013;
Matches 68; Conservative 39; Mismatches 102; Indels 69; Gaps 15;
QY 27 INNLYVNTIGGGEYQKYGAFDLKQKLYPEITFKQV-STGVADIPSQKITEWSAK 85
DB 106 VGDDKYFNPINGG--AASIGETIID--DKNY--YFQSGVLQTVF-----STEDGFK 153
QY 86 YFNGTNIL-----HRGSGYVLKADGGQYVNLGTTTKQF 118
DB 154 YFAPANTLDNLEGEAIDFTGKLIIDENIYFDDNYRGAVENKELGEMHYFSPETGKAF 213
QY 119 LPIQLTGEKKQ-----GNEGFKVGHGNYFYDLAGNMVKNFTIEDSYGNWYFFDQDG 171

; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-01763
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2366 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-480-604A-10

Query Match 8.1%; Score 115.5; DB 1; Length 2366;
Best Local Similarity 24.5%; Pred. No. 0.077;
Matches 68; Conservative 39; Mismatches 102; Indels 69; Gaps 15;

QY 27 INNLLVYVNTIGSGEYOKYKGGAFDLKQKLYPEIFTKQV-STGVAIDPSOKITEWSAK 85
Db 1859 VGDDKYFNPING--AASIGETIID--DKNY--YFNQSGVLQTVF-----STEDGFK 1906
QY 86 YFNGTNI-----HRSYVVKADGGYYNLGTTKQF 118
Db 1907 YFAPANTLENLEGEAIDFTGKLIIDENIYFDDNRYGAVENKELDGMHYPSPETGKAF 1966
QY 119 LPQLTGEKKQ-----GNEGFKGNDGNYFYDLAGNVMKNTFIEDSVGNWYFDDQG 171
Db 1967 KGLNQIGDYKYFNSDGMQKGFVSINDKNHYFDDSGVMKVGYTEID---GKHFFYAENG 2023
QY 172 KM---VEN-----KHFFV--DVDSYGEKGYFFFLKNVSGFRGLVQTDNGYYF-DNYGKM 220
Db 2024 EMQIGVENTEDGKFAHNEGLGNEEG-----EETSY-SGILFNFNKIYFDDSDFTAV 2076
QY 221 VRNOTINAGAMIVTLDNGK-LIKASVNSDAEYPTSTD 257
Db 2077 VGWKDLEDGSKYFDEDTAEAYIGLSINDGQYFND 2114

RESULT 24
US-08-405-496A-10
; Sequence 10, Application US/08405496A
; Patent No. 5919665
; GENERAL INFORMATION:

; APPLICANT: WILLIAMS, JAMES A.
; TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/405,496A
; FILING DATE: 16-MAR-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907

; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-01308
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2366 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-405-496A-10

Query Match 8.1%; Score 115.5; DB 2; Length 2366;
Best Local Similarity 24.5%; Pred. No. 0.077;
Matches 68; Conservative 39; Mismatches 102; Indels 69; Gaps 15;

QY 27 INNLLVYVNTIGSGEYOKYKGGAFDLKQKLYPEIFTKQV-STGVAIDPSOKITEWSAK 85
Db 1859 VGDDKYFNPING--AASIGETIID--DKNY--YFNQSGVLQTVF-----STEDGFK 1906
QY 86 YFNGTNI-----HRSYVVKADGGYYNLGTTKQF 118
Db 1907 YFAPANTLENLEGEAIDFTGKLIIDENIYFDDNRYGAVENKELDGMHYPSPETGKAF 1966
QY 119 LPQLTGEKKQ-----GNEGFKGNDGNYFYDLAGNVMKNTFIEDSVGNWYFDDQG 171
Db 1967 KGLNQIGDYKYFNSDGMQKGFVSINDKNHYFDDSGVMKVGYTEID---GKHFFYAENG 2023
QY 172 KM---VEN-----KHFFV--DVDSYGEKGYFFFLKNVSGFRGLVQTDNGYYF-DNYGKM 220
Db 2024 EMQIGVENTEDGKFAHNEGLGNEEG-----EETSY-SGILFNFNKIYFDDSDFTAV 2076
QY 221 VRNOTINAGAMIVTLDNGK-LIKASVNSDAEYPTSTD 257
Db 2077 VGWKDLEDGSKYFDEDTAEAYIGLSINDGQYFND 2114

RESULT 25
US-08-915-136-10
; Sequence 10, Application US/08915136
; Patent No. 6290960
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALLEY, BRUCE S.
; APPLICANT: PADHYE, NISHA V.
; APPLICANT: FIRCA, JOSEPH R.
; APPLICANT: STAFFORD, DOUGLAS C.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; PREVENTION OF C. DIFFICILE DISEASE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30

;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/915,136
;; FILING DATE:
;; CLASSIFICATION:
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/480,604
;; FILING DATE:
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/405,496
;; FILING DATE: 16-MAR-1995
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/329,154
;; FILING DATE: 25-OCT-1994
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/161,907
;; FILING DATE: 02-DEC-1993
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/985,321
;; FILING DATE: 04-DEC-1992
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/429,791
;; FILING DATE: 31-OCT-1989
;; ATTORNEY/AGENT INFORMATION:
;; NAME: INGOLIA, DIANE E.
;; REGISTRATION NUMBER: 40,027
;; REFERENCE/DOCKET NUMBER: OPHD-01763
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (415) 705-8410
;; TELEFAX: (415) 397-8338
;; INFORMATION FOR SEQ ID NO: 10:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 2366 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;;
;;
us-08-915-136-10

Query Match 8.1%; Score 115.5; DB 4: Length 2366;
Best Local Similarity 24.5%; Pred. No. 0.077;
Matches 68; Conservative 39; Mismatches 102; Indels 69; Gaps 15;

Qy 27 INNLYVYNTIGGGEYQKKGAFDLKQLYPEIFTKQV-STGVAIDPSQKITMSAK 85
Db 1859 VGDDKYFNPINGG--AASIGETIID--DKNY--YFQSGVLQTVF-----STEDGFK 1906

Qy 86 YFNGTNIL-----HRSGYVLKADGGQYINLCTTTKQF 118
Db 1907 YFAPANTLDENLEGEAIDFTGKLIIDENIYFDDNRYGAVEMKELDGMHYFSPETGKAF 1966

Qy 119 LPIQLTGKQKQ-----GNEGFKVKGNDGNYFYDLAGNMYKNTFIEDSVGNWYFDDQG 171
Db 1967 KGLAQIGDYKYFYFSDGVQKQGFVSINDKNHYFDDSGVMKVGTYEID---GKHFFYFENG 2023

Qy 172 KM-----KHVV--DVDSYGEKGYTFLLKNGVSFRGLVQTDNGTYF-DNYGKM 220
Db 2024 EMQIGVFNTPEDGKYFAHNNEDLNGEEG-----EEISY-SGILFNKNKIYFDDSTAV 2076

Qy 221 VRNQVINAGMIYTLDENGK-LIKASVNSDAEYPTSTD 257
Db 2077 VGWKDLEDGSKYFDEDTAEAYIGLSLINDQYIFNDD 2114

Search completed: August 12, 2002, 08:05:04
Job time: 146 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 12, 2002, 08:03:13 ; Search time 17.68 Seconds
(without alignments)
1451.123 Million cell updates/sec

Title: US-09-995-749A-2_COPY_1515_1781

Perfect score: 1430

Sequence: 1 IYNLPCKEVATVTRVDRGN.....SDABYPSTDVGMKLDQNK 267

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched:

283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

1: p1r1:**

2: p1r2:**

3: p1r3:**

4: p1r4:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	540.5	37.8	1508	2 T31098	probable dextran
2	539	37.7	1365	2 A41483	glucosyltransferase
3	510.5	35.7	1431	2 A45866	dextranucrase (EC
4	458.5	32.1	1475	2 B33135	gtfB protein precu
5	448	31.3	1449	2 T30857	glucosyltransferase
6	446	31.2	1449	2 T30552	glucosyltransferase
7	445.5	31.2	1375	2 J70345	dextranucrase (EC
8	439.5	30.7	1592	2 A38175	glucosyltransferase
9	439	30.7	1290	2 JCS473	dextranucrase (EC
10	437	30.6	1599	2 S22737	glucosyltransferase
11	428.5	30.0	1518	2 A44811	glucosyltransferase
12	418	29.2	1577	2 T30858	glucosyltransferase
13	169	11.8	2817	2 B97033	uncharacterized pr
14	156	10.9	316	1 A42936	N-acetylmutamoyl-L
15	154.5	10.8	563	2 A37184	glucan-binding pro
16	147.5	10.3	329	2 G97944	hypothetical prote
17	147.5	10.3	332	2 C95077	pneumococcal surfa
18	145.5	10.2	482	2 F86902	dextranucrase (EC
19	142	9.9	483	2 G86902	dextranucrase (EC
20	140.5	9.8	329	2 A55221	dextranase inhibito
21	136.5	9.5	601	2 S27962	cspC protein - Clo
22	135	9.4	318	1 S16016	N-acetylmutamoyl-L
23	132	9.2	318	2 A25634	N-acetylmutamoyl-L
24	132	9.2	318	2 D95226	autolysin (importe
25	132	9.2	318	2 H98090	N-acetylmutamoyl-L
26	131	9.2	316	1 A42935	N-acetylmutamoyl-L
27	129.5	9.1	338	2 A97914	choline-binding pr
28	129	9.0	721	2 C97980	endo-beta-N-acetyl
29	128.5	9.0	340	2 G95043	choline binding pr

30	128.5	9.0	658	2 E95111	endo-beta-N-acetyl
31	128	9.0	583	2 S57721	cspB protein - Clo
32	125	8.7	627	2 E95107	choline binding pr
33	125	8.7	2710	2 A37052	toxin A - Clostrid
34	124.5	8.7	648	2 S10869	enterotoxin A - C1
35	123.5	8.6	621	2 A95250	choline binding pr
36	123	8.6	627	2 G97975	hypothetical prote
37	123	8.6	2334	2 S32920	cell wall-associat
38	120.5	8.4	2364	2 I40884	cytotoxin L - Clo
39	116.5	8.1	590	2 I40608	cspA protein - Clo
40	116.5	8.1	690	2 F98114	choline-binding pr
41	116.5	8.1	2367	2 S70172	toxin B - Clostrid
42	115.5	8.1	635	2 S57714	cspB protein - Clo
43	115.5	8.1	2366	2 S10317	toxin B - Clostrid
44	115	8.0	2399	2 H71879	toxin-like outer m
45	114.5	8.0	339	1 M0BPC9	N-acetylmuramoyl-L
46	114.5	8.0	701	2 H98120	choline binding pr
47	113.5	7.9	2178	2 S55805	alpha-toxin - Clo
48	113	7.9	261	2 C75178	hypothetical prote
49	112	7.8	332	2 H95043	choline binding pr
50	112	7.8	501	2 F98050	lysozyme (EC 3.2.1
51	110.5	7.7	339	1 M0BPCP	N-acetylmuramoyl-L
52	109.5	7.7	1315	2 T28679	fibrinogen-binding
53	109.5	7.7	1385	2 D89824	hypothetical prote
54	107.5	7.5	1032	2 T17612	vp260 protein - Ch
55	106.5	7.4	340	2 D95045	choline binding pr
56	106.5	7.4	609	1 A42358	vibriolysin (EC 3.
57	106.5	7.4	762	2 A34355	cell surface prote
58	106.5	7.4	1296	2 I40645	botulinum neurotox
59	105	7.3	720	2 AF0242	probable exported
60	103.5	7.2	13055	2 T16580	hypothetical prote
61	103	7.2	693	2 H95255	choline binding pr
62	102.5	7.2	450	2 B95561	hypothetical prote
63	102	7.1	448	2 C95257	choline binding pr
64	102	7.1	676	2 H96970	endo-arabinase rel
65	102	7.1	1215	2 S60904	hypothetical prote
66	101.5	7.1	294	2 G97915	choline-binding pr
67	101.5	7.1	1962	2 A32634	lactocarpin (EC 3.4
68	101	7.1	308	2 F97037	hypothetical prote
69	101	7.1	448	2 C98122	choline binding pr
70	100.5	7.0	211	2 A95008	choline binding pr
71	100.5	7.0	253	2 A90558	hypothetical prote
72	100	7.0	554	2 A70120	hypothetical prote
73	100	7.0	1481	2 S28669	amylopullulanase p
74	99.5	7.0	1802	2 S28669	lactocarpin (EC 3.4
75	99.5	7.0	1902	2 S06997	lactocarpin (EC 3.4
76	98.5	6.9	1327	2 T14594	guanidine nucleoti
77	98	6.9	263	2 A83956	flagellar hook pro
78	98	6.9	285	2 C95045	choline binding pr
79	97.5	6.8	524	2 AG1455	oligopeptide ABC t
80	97.5	6.8	524	2 AH1091	oligopeptide ABC t
81	97.5	6.8	561	2 AH0452	hemolysin activato
82	97.5	6.8	1048	2 H64459	hypothetical prote
83	97	6.8	326	2 C91032	hypothetical prote
84	97	6.8	448	2 C81182	long-chain fatty a
85	97	6.8	448	2 D85876	long-chain fatty a
86	97	6.8	1024	2 S10056	hemolysin A - Esch
87	96.5	6.7	622	2 G86371	hypothetical prote
88	96.5	6.7	774	2 AG1565	autolysin (amidase
89	96	6.7	867	1 C64785	outer membrane ush
90	96	6.7	1302	2 C81182	iron-regulated pro
91	95.5	6.7	744	2 F95013	pneumococcal surfa
92	95.5	6.7	745	2 D82568	conserved hypotet
93	95.5	6.7	1651	2 JCI340	outer membrane pro
94	95.5	6.7	1653	2 E97835	hypothetical prote
95	95.5	6.7	1806	2 AF1717	probable peptidogl
96	95	6.6	653	2 F85620	partial fibrinogl u
97	95	6.6	653	2 H90756	partial fibrinogl u
98	95	6.6	1162	2 A97708	progenitor toxin n
99	95	6.6	1963	2 G86643	hypothetical prote
100	95	6.6	2167	2 AF1489	cell wall-associat

ALIGNMENTS

```
RESULT 1
T31098
probable dextranucrase (EC 2.4.1.5), extracellular - Leuconostoc mesenteroides
C:Species: Leuconostoc mesenteroides
C>Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 11-May-2000
C:Accession: T31098
R:Monchois, V.; Remaud-simeon, M.; Monsan, P.; Willemot, R.M.
FEBS Microbiol. Lett. 159, 307-315, 1998
A:Title: Cloning and sequencing of a gene coding for an extracellular dextranucrase (DS)
A:Reference number: Z20981; MUID:98164374
A:Accession: T31098
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1508 <MON>
A:Cross-references: EMBL:AF030129; NID:g2766611; PID:g2766612; PIDN:AA895453.1
A:Experimental source: strain NRRL B-1299
C:Genetics:
A:Gene: dsrB
C:Function:
A:Description: produces dextran encoding only of alpha(1-6) glucosidic bonds
C:Keywords: glycosyltransferase; hexosyltransferase

Query Match 37.8%; Score 540.5; DB 2; Length 1508;
Best Local Similarity 43.0%; Pred. No. 1.7e-32;
Matches 119; Conservative 47; Mismatches 76; Indels 35; Gaps 10;

QY 1 IYNLPKEVATVTRVDDRGVWVKDAIINNLYVNTIGGGEYQKYGAFDLKQLKLYPE 60
Db 1011 IYNLPQGLATVTRVNSYGDNDSDQSLYVQSGEGKYQAQYGGAFSLDIKKYPA 1070

QY 61 IFTKKQVSTGVAIDPSQKITEWSAKYFNGNIIHRSGSYVLKADG-QYYNL--GTTTKQ 117
Db 1071 LFTKQISTGLPMDPSQKITEWSKYPNGSNIQKAGYVLKDSGTQYKVKVSNNNNRD 1130

QY 118 FLPTQLGKKGQNEGVKNGDGNYYFYDLAGNMVKNFTIEDSVGNWYFFDQDKMYENK 177
Db 1131 FLPLQLDDLSE--TGVRNIGMVI-YTLGSLARMTFIQDNGNIIYFDSTGHLVTG- 1186

QY 178 HFVDVDSYGEKGYFFLKNGVSRFGGLVQ-TDNGTYFYFNDYGMVNRQTI----- 226
Db 1187 -FQINNH---HYFFLPNGELVQSFQNLQADGSTIYFDQKGRVFNQYTDQTGTAYTF 1241

QY 227 -NAGMIYT-----ID-----ENGLIKASYNSDAE 251
Db 1242 QNDGTMVTSGETIDGHKQYFYKNGTQVKGQFVSDTD 1278

RESULT 2
A41483
glucosyltransferase (EC 2.4.1.-) gtfS precursor - Streptococcus sobrinus
C:Species: Streptococcus sobrinus
C>Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 15-Oct-1999
C:Accession: A41483
R:Gilmore, K.S.; Russell, R.R.B.; Ferretti, J.J.
Infect. Immun. 58, 2452-2458, 1990
A:Title: Analysis of the Streptococcus downei gtfS gene, which specifies a glucosyltrans
A:Reference number: A41483; MUID:90316665
A:Accession: A41483
A:Molecule type: DNA
A:Residues: 1-1365 <GIL>
A:Cross-references: GB:M30943; NID:g153652; PIDN:AAA26898.1; PID:g153653
C:Genetics:
A:Gene: gtfS
C:Superfamily: cpl repeat homology
C:Keywords: glycosyltransferase; hexosyltransferase

Query Match 37.7%; Score 539; DB 2; Length 1365;
Best Local Similarity 42.1%; Pred. No. 1.9e-32;
```

```
Matches 131; Conservative 32; Mismatches 96; Indels 52; Gaps 11;

QY 1 IYNLPKEVATVTRVDDRGVWVKDAIINNLYVNTIGGGEYQKYGAFDLKQLKLYPE 60
Db 921 LYNLPKEVATVTRVDDRGVWVKDAIINNLYVNTIGGGEYQKYGAFDLKQLKLYPR 980

QY 61 IFTKKQVSTGVAIDPSQKITEWSAKYFNGNIIHRSGSYVLKADGQYYNLGTTTKQFLP 120
Db 981 LFTKQISTGLPMDPSQKITEWSAKYFNGSNIQKAGYVL-SEGKYLNL-ADGKLELP 1038

QY 121 IQLT---GEKKQNEGVKNGDGNYYFYDLAGNMVKNFTIEDSVGNWYFFDQDKMYENK 177
Db 1039 TVLNNYGPQPSANGFISKNGIHLV-DKNGQFVKNRKEIS-GSWYFFDSGKMATCK 1096

QY 178 HFVDVDSY-----GEKKGYFFLKNG-----VSFRG----- 202
Db 1097 TRIGNDTYLTFMPNGKOLKBEVWYDGKK-AYYDDNGRTWTRGFEVRVDGQDKMYENK 1155

QY 203 -----GLVQTDNGTYFYFNDYGMVNRQTIINAGAMITLTD-ENGLIKASYNSDAEYPTS 255
Db 1156 DGTIALGLVSLDRTLYFDAYGVQVAGQVITNGKSYTFDADQGLVQTDNANPA--PQG 1213

QY 256 TDVGRKMLDQNK 266
Db 1214 QAGWKLLGDNQ 1224

RESULT 3
A45866
dextranucrase (EC 2.4.1.5) precursor - Streptococcus mutans
C:Species: Streptococcus mutans
C>Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 07-May-1999
C:Accession: A45866
R:Honda, O.; Kato, C.; Kuramitsu, H.K.
J. Gen. Microbiol. 136, 2099-2105, 1990
A:Title: Nucleotide sequence of the Streptococcus mutans gtfD gene encoding the gluc
A:Reference number: A45866; MUID:91100958
A:Accession: A45866
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1431 <HON>
A:Cross-references: GB:M29296
C:Superfamily: cpl repeat homology
C:Keywords: glycosyltransferase; hexosyltransferase
F:181-201/Domain: cpl repeat homology <CP1>
F:1127-1146/Domain: cpl repeat homology <CP2>
F:1192-1211/Domain: cpl repeat homology <CP3>
F:1257-1276/Domain: cpl repeat homology <CP4>
F:1277-1297/Domain: cpl repeat homology <CP5>
F:1321-1340/Domain: cpl repeat homology <CP8>
F:1341-1361/Domain: cpl repeat homology <CP6>
F:1385-1404/Domain: cpl repeat homology <CP7>
```

```
Query Match 35.7%; Score 510.5; DB 2; Length 1431;
Best Local Similarity 43.6%; Pred. No. 2.8e-30;
Matches 120; Conservative 33; Mismatches 79; Indels 43; Gaps 12;

QY 1 IYNLPKEVATVTRVDDRGVWVKDAIINNLYVNTIGGGEYQKYGAFDLKQLKLYP 59
Db 966 IYNLPKEVATVTRVDDRGVWVKDAIINNLYVNTIGGGEYQKYGAFDLKQLKLYP 1025

QY 60 EIFTKKQVSTGVAIDPSQKITEWSAKYFNGNIIHRSGSYVLKADGG-QYYNL-GTTTKQ 117
Db 1026 SIFNRQISNGKKIDPSEKITEWSAKYFNGNIIHRSGSYVLKADGG-QYYNL-GTTTKQ 117

QY 118 FLPTQLTGEKKQNEGVKNGDGNYYFYDLAGNMVKNFTIEDSVGNWYFFDQDKMYENK 176
Db 1084 YLPQMT--NKEASTGFV--NDGNGMTFTSTGYQAKNSFVQDAKGNWYFFDNGHMYG 1139

QY 177 KHFDVDSYGEKGYFFLKNGVSRFGGLVQ-----DNGTYFYFNDYGMVNRQTI 219
Db 1140 LQQLN----GE--VQVFLSNGVQLRESFLENADGSKNYFGLGNRYNSGYSFNDKSKWR 1193
```

QY 222 -----FNQTNAGAMITY-LDENGKLIKASYNDAE 251

```

J. Bacteriol. 173, 989-996, 1991
A:Title: Peptide sequences for sucrose splitting and glucan binding within Streptococcus
A:Reference number: A38175; MUID:91112322
A:Accession: A38175
A>Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-1592 <ABO>
A:Cross-references: GB:D90213; NID:g217032; PIDN:BAAL4241.1; PID:d1014946; PID:g2170
C:Superfamily: cpl repeat homology
F:1093-1112/Domain: cpl repeat homology <CP1>
F:1222-1241/Domain: cpl repeat homology <CP2>
F:1287-1306/Domain: cpl repeat homology <CP3>
F:1330-1351/Domain: cpl repeat homology <CP4>
F:1352-1371/Domain: cpl repeat homology <CP5>
F:1402-1420/Domain: cpl repeat homology <CP6>
F:1465-1484/Domain: cpl repeat homology <CP7>
F:1513-1532/Domain: cpl repeat homology <CP8>

Query Match 30.7%; Score 439.5; DB 2; Length 1592;
Best Local Similarity 43.9%; Pred. No. 7.2e-25;
Matches 101; Conservative 38; Mismatches 72; Indels 19; Gaps 9;

QY 1 IYNLPGEVATVTRVDDRGVWVKDAIINNLYVNT-IGGSEYQKKGAFDLKQKLYP 59
Db 932 MYTFKQEVTVTRTDKFKPIAGSQIHSLVTDKSGDDYQAKYGGAFDELKQYP 991
QY 60 EFTFKQVSTGVAIDPSOKITWESAKYFNGTNIHRSGSYVLKAD-GGQYVNLGTTTKQF 118
Db 992 ELFTKKQSTGQVIDPSVKIKQWSAKYFNGSNILRGADYVLSQVSNKYFNASDT-LF 1050
QY 119 LPQLTGKKGQNEGFVKGNDGNYFYD--LAGNVKNTFIEDSVGMWFFPDQDGMVEN 176
Db 1051 LPSSELGKVVESGIRY----DQKGYIYNSATGQDKASFIETG-GLNYIFKQDGTMTVG 1105
QY 177 KHFVDVDSYGEKGYFFLKGVSFSGGLVQTD--NGTYFDNYGKMRNQ 224
Db 1106 AQTN-----GANVFFLENGTALR-NTIYTAQGNSHYVANDGKRYENE 1148

RESULT 9
JC5473
dextranucrase (EC 2.4.1.5) - Leuconostoc mesenteroides
C:Species: Leuconostoc mesenteroides
C:Date: 07-Jul-1997 #sequence_revision 29-Aug-1997 #text_change 29-Aug-1997
C:Accession: JC5473
R:Monchois, V.; Willemot, R.M.; Remaud-Simeon, M.; Croux, C.; Monsan, P.
A:Title: Cloning and sequencing of a gene coding for a novel dextranucrase from Leu
A:Reference number: JC5473; MUID:97136686
A:Accession: JC5473
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1290 <MON>
A:Cross-references: GB:U38181
C:Comment: This enzyme catalyzes the transfer of D-glucopyranosyl units from sucrose
C:Genetics:
A:Gene: dsra
C:Keywords: glycosyltransferase; hexosyltransferase
F:78-870/Domain: catalytic #status predicted <CAT>
F:922-1290/Domain: glucan-binding #status predicted <GCB>

Query Match 30.7%; Score 439; DB 2; Length 1290;
Best Local Similarity 41.3%; Pred. No. 5.9e-25;
Matches 93; Conservative 41; Mismatches 77; Indels 14; Gaps 6;

QY 1 IYNLPGEVATVTRVDDRGVWVKDAIINNLYVNT-IGGSEYQKKGAFDLKQKLYP 60
Db 771 IYNLPDQLVTAITDQSGHTVGSVIDHWLYASKTVAGGIYQQQYGGAFLEQLKQYQP 830
QY 61 IFTKKQVSTGVAIDPSOKITWESAKYFNGTNIHRSGSYVLKADGGQYVNLGTTTKQF 120
Db 1106 AQTN-----GANVFFLENGTALR-NTIYTAQGNSHYVANDGKRYENE 1148

J. Bacteriol. 169, 4263-4270, 1987
A:Title: Sequence analysis of the gtfB gene from Streptococcus mutans.
A:Reference number: A3135; MUID:87308013
A:Accession: C3135
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-349 <SHI>
A:Cross-references: GB:M17361
C:Genetics:
A:Gene: gtfC
C:Function:
A:Description: catalyzes the syntheses of both water-soluble and water-insoluble glucans
C:Superfamily: cpl repeat homology
C:Keywords: duplication; glycosyltransferase; hexosyltransferase
F:1-34/Domain: signal sequence #status predicted <SIG>
F:35-1375/Product: glucosyltransferase #status predicted <MAT>
F:1126-1145/Domain: cpl repeat homology <CP1>
F:1253-1272/Domain: cpl repeat homology <CP2>
F:1318-1337/Domain: cpl repeat homology <CP3>

Query Match 31.2%; Score 445.5; DB 2; Length 1375;
Best Local Similarity 39.8%; Pred. No. 2.1e-25;
Matches 115; Conservative 32; Mismatches 96; Indels 45; Gaps 13;

QY 1 IYNLPGEVATVTRVDDRGVWVKDAIINNLYVNT-IGGSEYQKKGAFDLKQKLYP 59
Db 961 MYALPEKQVSTGVAIDPSOKITWESAKYFNGTNIHRSGSYVLKADGGQYVNLGTTTKQF 1020
QY 60 EFTFKQVSTGVAIDPSOKITWESAKYFNGTNIHRSGSYVLKADGGQYVNLGTTTKQF 119
Db 1021 ELFAKQISTGVDPDPSVKIKQWSAKYFNGSNILRGADYVLSQVSNKYFNASDT-LF 1080
QY 120 PIQLTGKKGQNEGFVKG--NDG-NYFYDLAGNVKNTFIEDSVGMWFFPDQDGMVE 175
Db 1081 PKSLV-NPNHGTSSSYGLVFDGKGYIYNSATGQDKASFIETG-GLNYIFKQDGTMTVG 1137
QY 176 NKHFVDVDSYGEKGYFFLKGVSFSGGLVQTD--NGTYFDNYGKMRNQ 224
Db 1138 GAQGIN-----GANVFFLENGTALR-NTIYTAQGNSHYVANDGKRYENE 1148
QY 214 FDNVCKVNRQNTINAGMIYFLDENGKLIKASYNDAFYPTSTDVGMK 261
Db 1192 KQN-GIMAVGLTRVHGAQVY-FDASG-----FOAKGQFIITAD-GKL 1230

RESULT 8
A38175
glucosyltransferase precursor - Streptococcus sobrinus
C:Species: Streptococcus sobrinus
C:Date: 28-Aug-1992 #sequence_revision 28-Aug-1992 #text_change 15-Oct-1999
C:Accession: A38175
R:Abo, H.; Mtsamura, T.; Kodama, T.; Ohta, H.; Fukui, K.; Kato, K.; Kagawa, H.
```

Db 831 LFQKQKQISTQDPWNPDIQKSWKAYFNQSGNSIQGRGAWTVLKMGTQQYFVNSDAQTFELP 890
QY 121 IQLTGKQKQGNQGVF-KGNDGNYFYDLAGNWKNTFIEDSVGNWYFFDQDGKMYENKHF 179
Db 891 KQLLGEK--AKTGFYTRGKETS--FYSTSGYQAKSAFICDN-GNMYIEDDKGMYVGQV 945
QY 180 VVDSYGEKG-TYFFLKNQVSFRGLVQTDNGTYFYFDNYGKMYRN 223
Db 946 IN-----GINYFELPNGIELQDAYLVHDGMYYYNNITGKQLHN 983

RESULT 10
S22737
glucosyltransferase (EC 2.4.1.-) s - Streptococcus salivarius
C:Species: Streptococcus salivarius
C:Date: 16-Sep-1992 #sequence_revision 16-Sep-1992 #text_change 21-Jan-2000
C:Accession: S22737; S28810; B44811; S22727
R:Jacques, N.
submitted to the EMBL Data Library, March 1992
A:Reference number: S22726
A:Accession: S22737
A:Molecule type: DNA
A:Residues: 1-1599 <JAC>
A:Cross-references: EMBL:Z11872; NID:q47530; PIDN:CAA77898.1; PID:q47531
A:Experimental source: ATCC 25975
R:Giffard, P.M.; Simpson, C.L.; Milward, C.P.; Jacques, N.A.
J. Gen. Microbiol. 137, 2577-2593, 1991
A:Title: Molecular characterization of a cluster of at least two glucosyltransferase genes
A:Reference number: A44811; MUID:92148377
A:Accession: S28810
A:Molecule type: DNA
A:Residues: 1-51 <GIF>
A:Cross-references: EMBL:Z11873
C:Genetics:
A:Gene: gtfk
A:Superfamily: cpl repeat homology
C:Keywords: glycosyltransferase; hexosyltransferase
F:1456-1475/Domain: cpl repeat homology <CPR>

Query Match 30.6%; Score 437; DB 2; Length 1599;
Best Local Similarity 43.0%; Pred. No. 1.1e-24;
Matches 105; Conservative 28; Mismatches 91; Indels 20; Gaps 9;

QY 1 IYNLPKEVATVTRVDDRGVNWKDAIINNLYVYNT-IGGGEYQKKGAFDLKLOKLYP 59
Db 965 LYTLPGKEVYATRTIDTHGKVLDDISLVNKLXYTNTKSSGNDFAQYGGAFDLKLOKLYP 1024
QY 60 EIFTKKQVSTGVAIDPSOKITEWSAKYFNGTNIHRSGSYVLKADGQYVNL--GTTTK 116
Db 1025 EITKEVMEASGKIDFSVKLKQWEAKYFNGTNIHRSGSYVLKADGQYVNL--GTTTK 1080
QY 117 QFLPQTQTKGKQGNQGVFVKGNDGNYFYDLAGNWKNTFIEDSVGNWYFFDQDGKMYRN 176
Db 1081 -FLPAALTGDTK-AKTGFAYDGTGVY-YTTSQTQAKSQPVYTN-GKQYFNDKGVLVTG 1136
QY 177 KHFVDVDSYGEKGTFFLKNQVSFRGLVQTDNGTYFYFDNYGKMYRNQNTINAGAMITLD 236
Db 1137 EQTID-----GSNYFFLPNGVMTDGVNRNKGQSLV--YGRSGKLTTQTGKWEVTVKD 1188
QY 237 ENGK 240
Db 1189 DSGK 1192

RESULT 11
A44811
glucosyltransferase (EC 2.4.1.-) I - Streptococcus salivarius
C:Species: Streptococcus salivarius
C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 15-Oct-1999
C:Accession: A44811; S22726; S28809
R:Giffard, P.M.; Simpson, C.L.; Milward, C.P.; Jacques, N.A.
J. Gen. Microbiol. 137, 2577-2593, 1991

A:Title: Molecular characterization of a cluster of at least two glucosyltransferase
A:Reference number: A44811; MUID:92148377
A:Accession: A44811
A:Molecule type: DNA
A:Residues: 1-1518 <GIF>
A:Cross-references: EMBL:Z11873; NID:q47526; PIDN:CAA77900.1; PID:q47527
A:Note: sequence extracted from NCBI backbone (NCBIN:81050, NCBIP:81052)
C:Genetics:
A:Gene: gtfu
A:Superfamily: cpl repeat homology
C:Keywords: glycosyltransferase; hexosyltransferase
F:1307-1326/Domain: cpl repeat homology <CP4>

Query Match 30.0%; Score 428.5; DB 2; Length 1518;
Best Local Similarity 34.3%; Pred. No. 4.5e-24;
Matches 109; Conservative 42; Mismatches 80; Indels 87; Gaps 12;

QY 1 IYNLPKEVATVTRVDDRGVNWKDAIINNLYVYNT-IGGGEYQKKGAFDLKLOKLYP 59
Db 1005 IYQLPKEVYATRTIDTGAGKKTADAIIDHSLVANSKSGDYQAKYGGFELAEUKAYP 1064
QY 60 EIFTKKQVSTGVAIDPSOKITEWSAKYFNGTNIHRSGSYVLKADG-QQYVNLGTTTKQ- 117
Db 1065 EMFKYMNISTGKPIDDSVKLKQWEAKYFNGTNIHRSGSYVLKADGQYVNLGTTTKQ- 1121
QY 118 -FLPQLGEGEK-----QNGEYFVKGNDGNYFYDLAGNMY----- 152
Db 1122 NEIPLQLTKERKVIITGFFSGDGKGTFFYFUTSGTQAKSAFYTFN-GNTYTFDARGHMTNSE 1180
QY 153 -----KNTFIEDSVGNWYFF-----DQDG--K 172
Db 1181 YSPNGKDYVRFPLNGIMLSNAYIIDANGTYLYNSKGQMKGYTKFDVSETDKDGKSK 1240
QY 173 MYENKHFVD-----VDSYGEKGTFFLKNQVSFRGLVQTDNGTYFYDNYGKMY 221
Db 1241 VVKRYFTNEGVMKGVTVIDGF---TQYFGEDGQAKDKLVTFKGTYYFYDAHTGNGI 1296
QY 222 RNQTNAGAMITLDENG 239
Db 1297 KDTWRNKGWYFYFDANG 1314

RESULT 12
T30858
glucosyltransferase - Streptococcus salivarius
C:Species: Streptococcus salivarius
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
C:Accession: T30858
R:Simpson, C.L.; Giffard, P.M.; Jacques, N.A.
Infect. Immun. 63, 609-621, 1995
A:Title: Streptococcus salivarius ATCC 25975 possesses at least two genes coding for
A:Reference number: Z20909; MUID:95122197
A:Accession: T30858
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1577 <SIM>
A:Cross-references: EMBL:L35928; NID:q662380; PID:q662381; PIDN:AAC41413.1
C:Genetics:
A:Gene: gtfm

Query Match 29.2%; Score 418; DB 2; Length 1577;
Best Local Similarity 39.0%; Pred. No. 2.9e-23;
Matches 105; Conservative 38; Mismatches 82; Indels 44; Gaps 11;

QY 1 IYNLPKEVATVTRVDDRGVNWKDAIINNLYVYNT-IGGGEYQKKGAFDLKLOKLYP 59
Db 1042 IYNLPKEVYATRTNSYGTTPRPAEYLSLYAAKTFTGNDFGQYGGAFDLKAKYP 1101
QY 60 EIFTKKQVSTGVAIDPSOKITEWSAKYFNGTNIHRSGSYVLKADG-QQYVNL--GTTTK 116
Db 1102 AIFERVQISNGRKLITWNEKITQNSAKYFNGTNIHRSGSYVLKADGQYVNL--GTTTK 1159


```
Db 296 GKLSTGNTNLSQKKGKSFDP-----DTGEA-----WTNRFVN 332
QY 89 GTNHLHRGSGYV-----LKADGGYYMLGTTTKQFLP-----IQLTGEK 127
Db 333 AKYFFNFAGYVSTDFWYAGADGIGVTDKQIDGMDY-----FEPSSGIGVKGD- 383
QY 128 KQGNFVKGNNGNYFDL-AGNVKNTS-----IEDSVGNWYFDDG----- 171
Db 384 -----IAERDGVYLDSDSGVVKNRGTTPAERISTVIEARFPKTYFFGADGSKDL 436
QY 172 ---KMBNK---HFVDVDSYGEKGYFFLKKNGVSFRGGLVQTDNGTYFFDNYGKVRN--- 223
Db 437 TGMQIDGKYFFKDDHSIAKSEYSQIGGVP-DGGAEDGDGDFDQGFVFNREV 495
QY 224 -----QTINAGAMIYLDL---NKLKASVNS---DABYPTSTD 257
Db 496 RKYDYSNIWYYSDDKRVSGWTFIDGKRYFSQDEKTKGRIKGPITIDGKEYTFDKD 555
QY 258 VGKMLDON 265
Db 556 SGEVINSN 563

RESULT 16
G97944
hypothetical protein spr0583 [imported] - Streptococcus pneumoniae (strain R6)
C:Species: Streptococcus pneumoniae
C>Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 22-Oct-2001
C:Accession: G97944
R:Hoskins, J.A.; Albord Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; E
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M
Y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
A>Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A97872; MUID:21429245; PMID:11544234
A:Accession: G97944
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-329 <KUR>
A:Cross-references: GB:AE007317; PIDN:AAK99387.1; PID:g15458163; GSPDB:GN00174
C:Genetics:
A:Gene: spr0583

Query Match 10.3%; Score 147.5; DB 2; Length 329;
Best Local Similarity 27.9%; Pred. No. 0.0087;
Matches 55; Conservative 24; Mismatches 69; Indels 49; Gaps 11;

QY 88 NGTNILHRGS-GYVLKAD---GGQYNNLGT-----TKQFLPQLTGERKQNGEFPVK 137
Db 41 NGYSDIHEGEPGYSAKLDHRDHDGACELKNAPKAFKAKQSTAIQI-NTSSATTSGWVK- 98
QY 138 NDGNYYFYDLAGNVKNTFIEDSVGNW---YFPDQDGKMYENKHFVDVDSYGEKGYFFL 194
Db 99 ODGAWYFYDNGNGLVNA-----WQGSYYLKADGKMAQSEWIYD-SSY---QAWYYL 146
QY 195 KNGVSPRGGVLQTDNGTYFFDNYGKVRNQT-----NAGAMIYTL 235
Db 147 KSDGSYAKNAWQ---GAYYLSKNGKMAQGEWYDSSYQAWYYLKSQGSYARNWQGNYYL 203
QY 236 DENGKLIKASYNDAEY 252
Db 204 KSDGKMAKGEWYDATY 220

RESULT 18
F86902
dextranucrase (EC 2.4.1.5) [imported] - Lactococcus lactis subsp. lactis (strain IL
N;Alternate names: glucosyltransferase-S
C:Species: Lactococcus lactis subsp. lactis
C>Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
C:Accession: F86902
R:Boletín, A.; Wincker, P.; Manger, S.; Jaillon, O.; Malarne, K.; Weissenbach, J.; E
Genome Res. 11, 731-753, 2001
A>Title: The complete genome sequence of the lactic acid bacterium Lactococcus lacti
A:Reference number: A86625; MUID:21235186; PMID:11337471
A:Accession: F86902
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-482 <STO>
A:Cross-references: GB:AE005176; PID:g12725289; PIDN:AAK06320.1; GSPDB:GN00146
C:Genetics:
A:Gene: pspa
C:Keywords: glycosyltransferase; hexosyltransferase

Query Match 10.2%; Score 145.5; DB 2; Length 482;
Best Local Similarity 26.6%; Pred. No. 0.002;
Matches 64; Conservative 38; Mismatches 74; Indels 65; Gaps 17;

QY 23 KDALINNN---LYVWNTIGGGEYKKGAFADKLOKLYPEITFKKQVSTGVAIDSQKI 79
Db 108 KNGIVNENGERYYIINGVLOKGOONINGSWYL-----FNR---TGVNMYGQKQD 154
QY 80 T--EWSAKYPNGTNIHRSGSYVLKA---DGGQYNNLGTGTFKQFLPTLTGERKQNGE 134
Db 155 SNGQW-----YLPDRSSGIMKKGQVNDAGNWY-----LFDRSSGIMKKG----- 193
QY 135 VKGND-GNYFFYDLAGNVKNTFIEDSVGNWYFDDGKMYENKHFVDVDSYGEKGYFF 193
R:Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
```

```
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapp
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morri
A>Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: C95077
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-332 <KUR>
A:Cross-references: GB:AE005672; PIDN:AAK74812.1; PID:g14972141; GSPDB:GN00164; TIGR
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: SP0667

Query Match 10.3%; Score 147.5; DB 2; Length 332;
Best Local Similarity 27.9%; Pred. No. 0.0088;
Matches 55; Conservative 24; Mismatches 69; Indels 49; Gaps 11;

QY 88 NGTNILHRGS-GYVLKAD---GGQYNNLGT-----TKQFLPQLTGERKQNGEFPVK 137
Db 41 NGYSDIHEGEPGYSAKLDHRDHDGACELKNAPKAFKAKQSTAIQI-NTSSATTSGWVK- 98
QY 138 NDGNYYFYDLAGNVKNTFIEDSVGNW---YFPDQDGKMYENKHFVDVDSYGEKGYFFL 194
Db 99 ODGAWYFYDNGNGLVNA-----WQGSYYLKADGKMAQSEWIYD-SSY---QAWYYL 146
QY 195 KNGVSPRGGVLQTDNGTYFFDNYGKVRNQT-----NAGAMIYTL 235
Db 147 KSDGSYAKNAWQ---GAYYLSKNGKMAQGEWYDSSYQAWYYLKSQGSYARNWQGNYYL 203
QY 236 DENGKLIKASYNDAEY 252
Db 204 KSDGKMAKGEWYDATY 220
```

Db 194 QVNDAGNWLFDSSGIMRKGOVNDG--CNWYLFDRSSGVMRKGOVNDAGN-----YLF 246
QY 194 LK-NGVFRGGLVQTD-NG-TYFDNYGKMV-----RNOTIN--AGAMIYITLDENGK 240
Db 247 NRSSGIMYGG--LQTDINGKTYFENNAGKMWGTGTLEKNTYKENTTSECYALLSAGN 304
QY 241 L 241
Db 305 I 305
RESULT 19
G86902
dextranase [EC 2.4.1.5] [imported] - Lactococcus lactis subsp. lactis (strain IL1403)
N:Alternate names: glucosyltransferase-S
C:Species: Lactococcus lactis subsp. lactis
C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
C:Accession: G86902
R:Polotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarme, K.; Weissenbach, J.; Ehrlich
genome Res. 11, 731-753, 2001
A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis s
A:Reference number: A86625; MUID:21235186; PMID:11337471
A:Accession: G86902
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-483 <STO>
A:Cross-references: GB:AF005176; PID:g12725290; PIDN:AAK06321.1; GSPDB:GN00146
A:Experimental source: strain IL1403
C:Genetics:
A:Gene: pspB
C:Keywords: glycosyltransferase; hexosyltransferase
Query Match 9.9%; Score 142; DB 2; Length 483;
Best Local Similarity 25.4%; Pred. No. 0.0037;
Matches 61; Conservative 44; Mismatches 73; Indels 62; Gaps 18;
QY 23 KDAIINN--LVVNTIGGGEYKKGAFDQKLYPEFTKKQVSTGVAIDPSQKI 79
Db 97 KNQVNEGRIYINGVLQKQNGNSWYL-----FNR---TGVVMYGCQK- 142
QY 80 TWSAKYFNGTILHRSGYVLRKA---DGSYYNLTGTTKQFLIQLTGKKGNGEGFV- 135
Db 143 -DINGQWY----LFDSSGIMRKGOVNDAGNWLFDSS-----GIMRKGOVNDAGNWL 193
QY 136 -----KG--ND-GNYFYDLGAWKWNFTIEDSVGNWYFDQDKVKNKHFDVDS 184
Db 194 DRSSGIMRKGOVNDAGNWLFDSSGIMRKGOVNDG--GNWYLFDRSSGVMRKGOVNDAGN 252
QY 185 YGEKGYTFLLK-NGVFRGGLVQTD-NG-TYFDNYGKM-----VRNQ-----TINAGAM 231
Db 253 W-----YLFNRSSGIMYGG--LQTDINGKTYFENNAGKMWGTGTLEKNTYKENTTSECYALLSAGN 305
RESULT 20
A55221
dextranase inhibitor precursor - Streptococcus sobrinus (strain UAB66, serotype g)
C:Species: Streptococcus sobrinus
C:Date: 23-Mar-1995 #sequence_revision 23-Mar-1995 #text_change 17-Mar-1999
C:Accession: A55221
R:Sun, J.W.; Wanda, S.Y.; Camilli, A.; Curtiss III, R.
J. Bacteriol. 176, 7213-7222, 1994
A:Title: Cloning and DNA sequencing of the dextranase inhibitor gene (dei) from Streptoc
A:Reference number: A55221; MUID:95050304
A:Accession: A55221
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-329 <SUN>
A:Cross-references: GB:I34406
C:Superfamily: cpl repeat homology
F:173-192/Domain: cpl repeat homology <Cpl>

Query Match 9.8%; Score 140.5; DB 2; Length 329;
Best Local Similarity 22.3%; Pred. No. 0.0029;
Matches 69; Conservative 33; Mismatches 110; Indels 97; Gaps 11;
QY 10 ATVTVRDGRGNWVK-----DATTNNLNVVNTIGGGEYKKGAFDQKLYPEFTKKQVSTGVAIDPSQKI 51
Db 44 STVVRADSAVYTKPADETSQTDQAPATAEGTATANNQOASANTADQAEQORDTANQ 103
QY 52 DKLOKLYPEFTKKQVSTGVAIDPSQKITMSAKYFNGTILHRSGYVLRKADGQ----- 107
Db 104 DKQAV-DQASQPEQVAT--AVDQVQNAAKSDANQVWSTDV--KDSHAVVSKDDAKSSD 158
QY 108 -----YVNLGTTTQFLPIQLTGKKGNGEGFVKNNGN-----YTFYDLG 149
Db 159 QAEQAGFTTG-----NDWYKQEDGNLAKGLQTINGQTFYFDTWG 201
QY 150 NMVKNFTIEDSVGNWYFDQDKVKNKHFDVDSVGBKGVFFELKNGVSF----- 200
Db 202 KQVKSANTID-GKYYFDQDGDGDKWKRFRICIDQDYRGVAPGSKVGIAWLYQADGVS 260
QY 201 -----RGLVQTDNGTY-YFD-NYGMKVRNQTINAGAMIYTL 235
Db 261 ASGLTNTPDGRTLMFTYNNHDEQVKGKLVNTDGSNRYFDLHTGDMRLNTSLVDSQKYN 320
QY 236 DENGKLIKA 244
Db 321 DENGKLIKA 329
RESULT 21
S57962
cspC protein - Clostridium acetobutylicum (fragment)
C:Species: Clostridium acetobutylicum
C:Date: 13-Jan-1996 #sequence_revision 19-Apr-1996 #text_change 23-Mar-2001
C:Accession: S57962
R:Sanchez-Beato, A.; Garcia, J.
submitted to the EMBL Data Library, July 1995
A:Description: Molecular characterization of a family of choline-binding proteins of
A:Reference number: S57714
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-601 <SAN>
A:Cross-references: EMBL:Z50033; NID:g899232; PIDN:CAA90347.1; PID:g899233
C:Superfamily: cpl repeat homology
F:533-552/Domain: cpl repeat homology <CP3>
F:553-572/Domain: cpl repeat homology <CP4>
Query Match 9.5%; Score 136.5; DB 2; Length 601;
Best Local Similarity 23.5%; Pred. No. 0.013; 98; Indels 47; Gaps 12;
Matches 58; Conservative 44; Mismatches 44; Indels 47; Gaps 12;
QY 7 KEVATVTRVDDR-----GNWVKDAIINNLLVNTIGGGEYKKGAFDQKLYPEFTKKQVSTGVAIDPSQKI 57
Db 392 KEDCEVSSQDDTSVQTDVQGNLNR--LDGVIYKFDNTDDWVKVKGDSF-DEL--- 445
QY 58 YPEFTKKQVSTGVAIDPSQKITMSAKYFNGTILHRSGYVLRKADGQYVNLGTTTQ 117
Db 446 --SVYDKNIAAASEDDVDVYSLIGKSSDNNNTPEPTTGWVQAADGTWYVYKEDGTKA 503
QY 118 FLPIQLTGKKGNGEGFVKNNGNRYFYFDLAGNVMKNFTIEDSVGNWYFDQDKVKNKH 177
Db 504 -----TGWLNIN-GVWYVYNTDGVVATGML--NLNGVWYILNPSGAMAT-S- 545
QY 178 HFVDVDSYGEKGYTFLLKNGVFRGGLVQTDNGT-YFDNYGKVRNQTINAGAMIYTL 236
Db 546 -WLNIN-----GTWYVYLNSSGAMATGML-NDNGTWYVYLNSSGAMLYNTVDG----YVLG 594
QY 237 ENKGLIK 243
Db 595 SNGAWIR 601

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 12, 2002, 08:05:08 ; Search time 10.33 Seconds
(without alignments)
1000.786 Million cell updates/sec

Title: US-09-995-749A-2_COPY_1515_1781

Perfect score: 1430

Sequence: 1 IYNLPGRVAVTRVDRGN.....SDAEYPISTDVGRMLDQNK 267

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	539	37.7	1365	1	GTF5_STRDO
2	510.5	35.7	1462	1	GTF5_STRMU
3	458.5	32.1	1476	1	GTF5_STRMU
4	445.5	31.2	1375	1	GTF5_STRMU
5	439.5	30.7	1592	1	GTF5_STRDO
6	432.5	30.2	1597	1	GTF5_STRDO
7	135	9.4	318	1	ALYS_BPHB3
8	132	9.2	318	1	ALYS_STRPN
9	128.5	9.0	658	1	LYTE_STRPN
10	125	8.7	2710	1	TOXA_CLODI
11	123	8.6	2334	1	WAPA_BACSU
12	115.5	8.1	2366	1	TOXB_CLODI
13	114.5	8.0	339	1	LYCA_BPCP9
14	110.5	7.7	339	1	LYCA_BPCP1
15	106.5	7.4	609	1	HAPT_VTBCH
16	106.5	7.4	762	1	SLAP_ACEKI
17	106.5	7.4	1295	1	BXA2_CLOBO
18	101.5	7.1	1902	1	P3P_IAGLC
19	100	7.0	1481	1	APUL_THRET
20	99.5	7.0	1902	1	P1P_IAGLC
21	99.5	7.0	1902	1	P2P_IAGLC
22	97.5	6.8	1705	1	CYAA_BORBR
23	97	6.8	1024	1	HLVA_ECOLI
24	96.5	6.7	639	1	BMPH_STRPU
25	96	6.7	867	1	SFMD_ECOLI
26	96	6.7	1302	1	FRPA_NEIMB
27	95.5	6.7	299	1	SM30_BOVIN
28	95.5	6.7	1655	1	OMPB_RICCN
29	95	6.6	1162	1	BXEN_CLOBO
30	95	6.6	1656	1	OMPB_RICJA
31	94.5	6.6	434	1	VG05_VARV
32	94.5	6.6	1295	1	BXA1_CLOBO
33	94	6.6	448	1	FADL_ECOLI

RESULT 1

ALIGNMENTS

34	94	6.6	1	YCBS_ECOLI	P75857	escherichia
35	93.5	6.5	1	CSG_METFE	P27373	methanother
36	93.5	6.5	1	CSG_METSC	P27374	methanother
37	93.5	6.5	1	CHAC_PEDHE	Q59288	pseudobacter
38	93.5	6.5	1	ISOA_PSEAY	P10342	pseudomonas
39	91.5	6.4	1	PRIM_UREPA	Q99266	ureaplasma
40	91.5	6.4	1	ISOA_PSESP	P25501	pseudomonas
41	91.5	6.4	1	HYSA_STRAU	Q59801	staphylococ
42	91.5	6.4	1	OMPB_RICNY	P96989	r outer mem
43	91	6.4	1	PMPA_CHLMU	Q99193	chlamydia m
44	90.5	6.3	1	HLV1_ECOLI	P09983	escherichia
45	90.5	6.3	1	P2P_LACPA	Q02470	lactobacilli
46	90.5	6.3	1	SMX1_SCHMA	Q02500	schistosoma
47	90	6.3	1	ADCA_STRPN	Q05703	streptococc
48	90	6.3	1	YAB3_MYCPN	P75610	mycoplasma
49	90	6.3	1	PMPG_CHLMU	Q99145	chlamydia m
50	90	6.3	1	OMPL_PROPR	Q52581	photobacter
51	89.5	6.3	1	YA22_AQUAE	O67134	aquifex seo
52	89.5	6.3	1	PM16_CHLPN	Q94882	chlamydia p
53	89.5	6.3	1	BIGA_SALTY	P25927	salmonella
54	89.5	6.3	1	NANH_CLOSO	P15698	clostridium
55	89	6.2	1	YIEC_ECOLI	P26218	escherichia
56	89	6.2	1	YKDA_MITCA	P45615	mycoplasma
57	89	6.2	1	HYSA_STRPN	Q54873	streptococc
58	89	6.2	1	BXEN_CLOBU	Q06366	clostridium
59	89	6.2	1	RPL1_PLAFA	P27625	plasmodium
60	89	6.2	1	FULA_THEMA	Q03840	thermotoga
61	88.5	6.2	1	STRH_STRPN	P49610	streptococc
62	88.5	6.2	1	RPAL_METH	Q27125	methanobact
63	88	6.2	1	YFJA_ECOLI	P52143	escherichia
64	87	6.1	1	PMPG_CHLTR	O84879	chlamydia t
65	87	6.1	1	SD30_RAT	Q03336	rattus norv
66	86.5	6.0	1	CSGT_BACS2	P31746	bacillus sp
67	86.5	6.0	1	SIF3_YEAST	P38717	saccharomyc
68	86.5	6.0	1	Y338_MYCGE	P47580	mycoplasma
69	86.5	6.0	1	CYAA_BORPE	P15318	borderella
70	86.5	6.0	1	PENK_MESAU	P50175	mesocricetu
71	86	6.0	1	PHOE_ECOLI	P02932	escherichia
72	86	6.0	1	GLPK_THEFL	O66131	thermus aqu
73	86	6.0	1	ATYM_BACST	P19531	bacillus st
74	86	6.0	1	BGAL_CLOAB	P24131	clostridium
75	86	6.0	1	AIDA_ECOLI	Q03155	escherichia
76	86	6.0	1	SM30_MOUSE	Q64374	mus musculus
77	85.5	6.0	1	UP05_ECOLI	P39170	escherichia
78	85.5	6.0	1	CS01_ECOLI	P53512	escherichia
79	85.5	6.0	1	YPX2_CAEEL	Q02056	caenorhabdi
80	85.5	6.0	1	MSH6_YEAST	Q03834	saccharomyc
81	85.5	6.0	1	FMKE_ECOLI	P02970	escherichia
82	85	5.9	1	GLPK_THEAQ	Q94X53	thermus aqu
83	85	5.9	1	RT03_ORNBE	P27754	oenothera b
84	85	5.9	1	LP1G_DROME	P11997	drosophila
85	85	5.9	1	ISOA_FLASP	O34611	flavobacter
86	85	5.9	1	HFC1_HAEIN	P33397	haemophilus
87	85	5.9	1	YDGH_RACSU	P96706	bacillus su
88	85	5.9	1	MOK1_SCHPO	Q94588	schizosacch
89	85	5.9	1	HAP_HAEIN	P45387	haemophilus
90	84.5	5.9	1	RO22_XENLA	P51990	xenopus lae
91	84	5.9	1	CHS3_CANAL	P16405	plasmodium
92	84	5.9	1	Y32K_SSV1	P30573	candida alb
93	84	5.9	1	SUBT_BACS9	P20197	sulfolobus
94	83.5	5.8	1	AURE_STAAU	P28842	bacillus ap
95	83.5	5.8	1	NPRV_VIBPR	P81177	staphylococ
96	83.5	5.8	1	YLS8_CAEEL	Q00971	vibrio prot
97	83.5	5.8	1	DPOL_HSV6U	P28857	human herpe
98	83.5	5.8	1	MRX4_DROME	Q94887	drosophila
99	83.5	5.8	1			
100	83.5	5.8	1			

```

981 LFSTKQISTGKPIDPSVKTWNRSKYFNGSNILGAKYVL-SGKNYLN-ADGKFLFP 1038
QY 121 IQLT---GEKKQGNESFVKGNDGNYFYFDLAGNMVKNFTIEDSYGNWYFFDQDGKHYENK 177
D 1039 TVLNNYTGQPVQSANGFISKNGGIHYL-DKNGQEVKRFKEIS-GSWYFDSGKNATGK 1096
QY 178 HFVDVDSY-----GEKGTYPFLNG-----VSFRG----- 202
D 1097 TKIGNDIYLFPMNGKOLKEGVWYDGGK-AYIDDNGRFTWKNKGFEVFGDQDKWYFNG 1155
QY 203 -----GLVQTQDNGTYFDNYKMYRNQNTINAGAMIYPLD-ENGLIKASYNDAEYPTS 255
D 1156 DGTIAIGVLSLNRNLTLYDAYGYQYVGQTVINGKSYTFDADGDIYQTDNANPA--PQ 1213
QY 256 TDVGKMLDNQK 266
D 1214 QAGWKELGDNQ 1224

RESULT 2
GTFFD_STRMU STANDARD: PRT: 1462 AA.
AC F49331; 069386; 069386; 069389; 069392; 069398;
DT 01-FEB-1996 (Rel. 33, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Glucosyltransferase-S precursor (EC 2.4.1.5) (GTF-S) (Dextranucrase)
DE (Sucrose 6-glucosyltransferase).
CN GTFD.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OC NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GS-5;
RA MEDLINE=91100958; PubMed=2148600;
RA Honda O., Kato C., Kuramitsu H.K.;
RT "Nucleotide sequence of the Streptococcus mutans gtfD gene encoding
RT the glucosyltransferase-S enzyme.";
RL J. Gen. Microbiol. 136:2099-2105(1990).
RP SEQUENCE FROM N.A.
RC STRAIN=MT4239, MT4245, MT4251, MT4467, AND MT8148;
RC MEDLINE=98231643; PubMed=9570124;
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,
RA Kimura S., Hamada S.;
RT "Molecular analyses of glucosyltransferase genes among strains of
RT Streptococcus mutans.";
RL FEMS Microbiol. Lett. 161:331-336(1998).
CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT
CC OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE
CC AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.
CC -1- CATALYTIC ACTIVITY: Sucrose + [(1,6)-alpha-D-glucosyl](N) - D-
CC fructose + [(1,6)-alpha-D-glucosyl](N+1).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- DISEASE: DENTAL CARIES.
CC -1- MISCELLANEOUS: GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA
CC 1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES
CC WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH
CC FORMS OF GLUCANS.
CC -1- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-
CC BINDING PROTEIN FROM S.MUTANS.

THIS SWISS-PROT entry is copyright. It is produced through a collabora-
between the Swiss Institute of Bioinformatics and the EMBL outstati-
the European Bioinformatics Institute. There are no restrictions on
use by non-profit institutions as long as its content is in no
modified and this statement is not removed. Usage by and for commer-
entities requires a license agreement (See http://www.isb-sib.ch/annou
or send an email to license@isb-sib.ch).

```


APPLICATION NUMBER: US 08/457,459
FILING DATE: 01-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Bak, Mary E.
REGISTRATION NUMBER: 31,215
REFERENCE/DOCKET NUMBER: WST49DUSA
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-540-9206
TELEFAX: 215-540-5818
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 71 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-555-678-24

Query Match 0.4%; Score 7; DB 1; Length 71;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 283 NGKEVAR 289
|||||||
DB 40 NGKEVAR 46

RESULT 14
PCT-US95-02275-24
Sequence 24, Application PC/WUS9502275
GENERAL INFORMATION:
APPLICANT: Wistar Institute of Anatomy & Biology
TITLE OF INVENTION: RNA Editing Enzyme and Methods of Use
NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESS:
ADDRESSEE: Howson and Howson
STREET: Spring House Corporate Cntr, P.O. Box 457
CITY: Spring House
STATE: Pennsylvania
COUNTRY: USA
ZIP: 19477
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/02275
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/280,443
FILING DATE: 25-JUL-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/197,794
FILING DATE: 17-FEB-1994
ATTORNEY/AGENT INFORMATION:
NAME: Bak, Mary E.
REGISTRATION NUMBER: 31,215
REFERENCE/DOCKET NUMBER: WST49BPCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-540-9206
TELEFAX: 215-540-5818
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 71 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: protein
PCT-US95-02275-24

Query Match 0.4%; Score 7; DB 5; Length 71;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 283 NGKEVAR 289
|||||||
DB 40 NGKEVAR 46

RESULT 15
US-08-436-420-34
Sequence 34, Application US/08436420
Patent No. 5840524
GENERAL INFORMATION:
APPLICANT: VAN DAMME, Jo; and
APPLICANT: PROOST, Paul
TITLE OF INVENTION: GRANULOCYTE CHEMOTACTIC PROTEIN
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: HALE and DORR LLP
STREET: 1455 PENNSYLVANIA AVENUE, N.W.
CITY: WASHINGTON
STATE: DISTRICT OF COLUMBIA
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/436,420
FILING DATE: 24-MAY-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/BF93/03330
FILING DATE: 26-NOV-1993
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/982,539
FILING DATE: 27-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: BAKER, Hollie L.
REGISTRATION NUMBER: 31,321
REFERENCE/DOCKET NUMBER: 102378.215
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 942-8400
TELEFAX: (202) 942-8484
INFORMATION FOR SEQ ID NO: 34:
SEQUENCE CHARACTERISTICS:
LENGTH: 81 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-436-420-34

Query Match 0.4%; Score 7; DB 2; Length 81;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 639 LVNGKEV 645
|||||||
DB 48 LVNGKEV 54

RESULT 16
US-09-214-095D-4
Sequence 4, Application US/09214095D
Patent No. 6280987

GENERAL INFORMATION:
; APPLICANT: Landry, Donald
; TITLE OF INVENTION: ANTI-COCAINE CATALYTIC ANTIBODY
; FILE REFERENCE: 51400-A-PCT-US
; CURRENT APPLICATION NUMBER: US/09/214, 095D
; CURRENT FILING DATE: 1999-07-19
; NUMBER OF SEQ ID NOS: 121
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 4
; LENGTH: 98
; TYPE: PRT
; ORGANISM: Murinae gen.sp.
US-09-214-095D-4

Query Match 0.4%; Score 7; DB 4; Length 98;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1007 GTTANN 1013
Db 15 GTTANN 21

RESULT 17
US-08-672-345C-94
; Sequence 94, Application US/08672345C
; Patent No. 5948658
; GENERAL INFORMATION:
; APPLICANT: Landry Donald, W.
; TITLE OF INVENTION: ANTI-COCAINE CATALYTIC ANTIBODY
; NUMBER OF SEQUENCES: 108
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper and Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/672,345C
; FILING DATE: 24-JUN-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 0575/51400
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-278-0400
; TELEFAX: 212-391-0525
; INFORMATION FOR SEQ ID NO: 94:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 99 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-672-345C-94

Query Match 0.4%; Score 7; DB 2; Length 99;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1007 GTTANN 1013
Db 16 GTTANN 22

RESULT 18
US-09-214-095D-96
; Sequence 96, Application US/09214095D
; Patent No. 6280987
; GENERAL INFORMATION:
; APPLICANT: Landry, Donald
; TITLE OF INVENTION: ANTI-COCAINE CATALYTIC ANTIBODY
; FILE REFERENCE: 51400-A-PCT-US
; CURRENT APPLICATION NUMBER: US/09/214, 095D
; CURRENT FILING DATE: 1999-07-19
; NUMBER OF SEQ ID NOS: 121
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 96
; LENGTH: 100
; TYPE: PRT
; ORGANISM: Murine
US-09-214-095D-96

Query Match 0.4%; Score 7; DB 4; Length 100;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1007 GTTANN 1013
Db 16 GTTANN 22

RESULT 19
US-08-672-345C-4
; Sequence 4, Application US/08672345C
; Patent No. 5948658
; GENERAL INFORMATION:
; APPLICANT: Landry Donald, W.
; TITLE OF INVENTION: ANTI-COCAINE CATALYTIC ANTIBODY
; NUMBER OF SEQUENCES: 108
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper and Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/672,345C
; FILING DATE: 24-JUN-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 0575/51400
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-278-0400
; TELEFAX: 212-391-0525
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 109 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-672-345C-4

Query Match 0.4%; Score 7; DB 2; Length 109;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1007 GTITANN 1013
   |||||
Db 26 GTITANN 32

RESULT 20
US-07-857-224B-7
; Sequence 7, Application US/07857224B
; Patent No. 5958784
; GENERAL INFORMATION:
; APPLICANT: Benner, Steven A.
; TITLE OF INVENTION: Predicting Folded Structures of Proteins
; NUMBER OF SEQUENCES: 114
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Steven A. Benner
; STREET: Hadlaubstrasse 151
; CITY: Zurich
; STATE: none
; COUNTRY: Switzerland
; ZIP: (note: this is an international post code) CH-8092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch diskette, 1.4 Mb storage
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 7.0
; SOFTWARE: Microsoft Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/857,224B
; FILING DATE: 03/25/92
; CLASSIFICATION: 436
; PRIOR APPLICATION DATA: none
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (international) 41 1 632 2830
; TELEFAX: (international) 41 1 262 2437
; TELEX: none
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 260
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE:
; DESCRIPTION: protein
; ORIGINAL SOURCE:
; ORGANISM: Saccharomyces cerevisiae
; FEATURE: Protein kinase, Table 8 Column 7
; PUBLICATION INFORMATION:
; AUTHORS:
; HANKS, S. K.
; QUINN, A. M.
; HUNTER, T.
; TITLE: The protein kinase family
; JOURNAL: Science
; VOLUME: 241
; PAGES: 42-52
; DATE: 1988
US-07-857-224B-7

Query Match 0.4%; Score 7; DB 2; Length 260;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 228 KSVDMWS 234
   |||||
Db 177 KSVDMWS 183

RESULT 21
US-07-857-224B-8
; Sequence 8, Application US/07857224B
; Patent No. 5958784
; GENERAL INFORMATION:
; APPLICANT: Benner, Steven A.

```

```

; TITLE OF INVENTION: Predicting Folded Structures of Proteins
; NUMBER OF SEQUENCES: 114
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Steven A. Benner
; STREET: Hadlaubstrasse 151
; CITY: Zurich
; STATE: none
; COUNTRY: Switzerland
; ZIP: (note: this is an international post code) CH-8092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch diskette, 1.4 Mb storage
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 7.0
; SOFTWARE: Microsoft Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/857,224B
; FILING DATE: 03/25/92
; CLASSIFICATION: 436
; PRIOR APPLICATION DATA: none
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (international) 41 1 632 2830
; TELEFAX: (international) 41 1 262 2437
; TELEX: none
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 260
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE:
; DESCRIPTION: protein
; ORIGINAL SOURCE:
; ORGANISM: Saccharomyces cerevisiae
; FEATURE: Protein kinase, Table 8 Column 8
; PUBLICATION INFORMATION:
; AUTHORS:
; HANKS, S. K.
; QUINN, A. M.
; HUNTER, T.
; TITLE: The protein kinase family
; JOURNAL: Science
; VOLUME: 241
; PAGES: 42-52
; DATE: 1988
US-07-857-224B-8

Query Match 0.4%; Score 7; DB 2; Length 260;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 228 KSVDMWS 234
   |||||
Db 177 KSVDMWS 183

RESULT 22
US-07-857-224B-13
; Sequence 13, Application US/07857224B
; Patent No. 5958784
; GENERAL INFORMATION:
; APPLICANT: Benner, Steven A.
; TITLE OF INVENTION: Predicting Folded Structures of Proteins
; NUMBER OF SEQUENCES: 114
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Steven A. Benner
; STREET: Hadlaubstrasse 151
; CITY: Zurich
; STATE: none
; COUNTRY: Switzerland
; ZIP: (note: this is an international post code) CH-8092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch diskette, 1.4 Mb storage
; COMPUTER: Apple Macintosh

```

```

; OPERATING SYSTEM: Macintosh 7.0
; SOFTWARE: Microsoft Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/857,224B
; FILING DATE: 03/25/92
; CLASSIFICATION: 436
; PRIOR APPLICATION DATA: none
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (International) 41 1 632 2830
; TELEFAX: (International) 41 1 262 2437
; TELEX: none
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 263
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE:
; DESCRIPTION: protein
; ORIGINAL SOURCE:
; ORGANISM: rabbit
; FEATURE: Protein kinase; Table 8 Column 14
; PUBLICATION INFORMATION:
; AUTHORS:
; AUTHORS: Hanks, S. K.
; AUTHORS: Quinn, A. M.
; AUTHORS: Hunter, T.
; TITLE: The protein kinase family
; JOURNAL: Science
; VOLUME: 241
; PAGES: 42-52
; DATE: 1988
; US-07-857-224B-13

```

```

Query Match 0.4%; Score 7; DB 2; Length 263;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 228 KSVDMWS 234
DB 180 KSVDMWS 186

RESULT 23
US-07-857-224B-10
; Sequence 10, Application US/07857224B
; Patent No. 5958784
; GENERAL INFORMATION:
; APPLICANT: Benner, Steven A.
; TITLE OF INVENTION: Predicting Folded Structures of Proteins
; NUMBER OF SEQUENCES: 114
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Steven A. Benner
; STREET: Hadlaubstrasse 151
; CITY: Zurich
; STATE: none
; COUNTRY: Switzerland
; ZIP: (note: this is an international post code) CH-8092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch diskette, 1.4 Mb storage
; OPERATING SYSTEM: Macintosh 7.0
; SOFTWARE: Microsoft Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/857,224B
; FILING DATE: 03/25/92
; CLASSIFICATION: 436
; PRIOR APPLICATION DATA: none
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (International) 41 1 632 2830
; TELEFAX: (International) 41 1 262 2437
; TELEX: none
; INFORMATION FOR SEQ ID NO: 10:

```

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 264
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE:
; DESCRIPTION: protein
; ORIGINAL SOURCE:
; ORGANISM: bovine
; FEATURE: Protein kinase; Table 8 Column 11
; PUBLICATION INFORMATION:
; AUTHORS:
; AUTHORS: Hanks, S. K.
; AUTHORS: Quinn, A. M.
; AUTHORS: Hunter, T.
; TITLE: The protein kinase family
; JOURNAL: Science
; VOLUME: 241
; PAGES: 42-52
; DATE: 1988
; US-07-857-224B-10

Query Match 0.4%; Score 7; DB 2; Length 264;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 227 GKSDWM 233
DB 180 GKSDWM 186

RESULT 24
US-07-857-224B-11
; Sequence 11, Application US/07857224B
; Patent No. 5958784
; GENERAL INFORMATION:
; APPLICANT: Benner, Steven A.
; TITLE OF INVENTION: Predicting Folded Structures of Proteins
; NUMBER OF SEQUENCES: 114
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Steven A. Benner
; STREET: Hadlaubstrasse 151
; CITY: Zurich
; STATE: none
; COUNTRY: Switzerland
; ZIP: (note: this is an international post code) CH-8092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch diskette, 1.4 Mb storage
; OPERATING SYSTEM: Macintosh 7.0
; SOFTWARE: Microsoft Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/857,224B
; FILING DATE: 03/25/92
; CLASSIFICATION: 436
; PRIOR APPLICATION DATA: none
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (International) 41 1 632 2830
; TELEFAX: (International) 41 1 262 2437
; TELEX: none
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 264
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE:
; DESCRIPTION: protein
; ORIGINAL SOURCE:
; ORGANISM: rabbit
; FEATURE: Protein kinase; Table 8 Column 12
; PUBLICATION INFORMATION:
; AUTHORS:
; AUTHORS: Hanks, S. K.

```

;
; AUTHORS: Quinn, A. M.
; TITL: The protein kinase family
; JOURNAL: Science
; VOLUME: 241
; PAGES: 42-52
; DATE: 1988
US-07-857-224B-11

Query Match 0.4%; Score 7; DB 2; Length 264;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 227 GKSVDWW 233
|111111|
DB 180 GKSVDWW 186

RESULT 25
US-07-857-224B-12
; Sequence 12, Application US/07857224B
; Patent No. 5958784
; GENERAL INFORMATION:
; APPLICANT: Benner, Steven A.
; TITLE OF INVENTION: Predicting Folded Structures of Proteins
; NUMBER OF SEQUENCES: 114
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Steven A. Benner
; STREET: Hadlaubstrasse 151
; CITY: Zurich
; STATE: none
; COUNTRY: Switzerland
; ZIP: (note: this is an international post code) CH-8092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch diskette, 1.4 Mb storage
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 7.0
; SOFTWARE: Microsoft Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/857,224B
; FILING DATE: 03/25/92
; CLASSIFICATION: 436
; PRIOR APPLICATION DATA: none
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (International) 41 1 632 2830
; TELEFAX: (International) 41 1 262 2437
; TELEX: none
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 264
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: bovine
; FEATURE: Protein kinase; Table 8 Column 13
; PUBLICATION INFORMATION:
; AUTHORS:
; AUTHORS: Hanks, S. K.
; AUTHORS: Quinn, A. M.
; AUTHORS: Hunter, T.
; TITLE: The protein kinase family
; JOURNAL: Science
; VOLUME: 241
; PAGES: 42-52
; DATE: 1988
US-07-857-224B-12

Query Match 0.4%; Score 7; DB 2; Length 264;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 227 GKSVDWW 233
|111111|
DB 180 GKSVDWW 186

RESULT 26
US-07-857-224B-15
; Sequence 15, Application US/07857224B
; Patent No. 5958784
; GENERAL INFORMATION:
; APPLICANT: Benner, Steven A.
; TITLE OF INVENTION: Predicting Folded Structures of Proteins
; NUMBER OF SEQUENCES: 114
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Steven A. Benner
; STREET: Hadlaubstrasse 151
; CITY: Zurich
; STATE: none
; COUNTRY: Switzerland
; ZIP: (note: this is an international post code) CH-8092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch diskette, 1.4 Mb storage
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 7.0
; SOFTWARE: Microsoft Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/857,224B
; FILING DATE: 03/25/92
; CLASSIFICATION: 436
; PRIOR APPLICATION DATA: none
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (International) 41 1 632 2830
; TELEFAX: (International) 41 1 262 2437
; TELEX: none
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 264
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; DESCRIPTION: protein
; ORIGINAL SOURCE:
; ORGANISM: bovine
; FEATURE: Protein kinase; Table 8 Column 16
; PUBLICATION INFORMATION:
; AUTHORS:
; AUTHORS: Hanks, S. K.
; AUTHORS: Quinn, A. M.
; AUTHORS: Hunter, T.
; TITLE: The protein kinase family
; JOURNAL: Science
; VOLUME: 241
; PAGES: 42-52
; DATE: 1988
US-07-857-224B-15

Query Match 0.4%; Score 7; DB 2; Length 264;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 227 GKSVDWW 233
|111111|
DB 180 GKSVDWW 186

RESULT 27
US-07-857-224B-17
; Sequence 17, Application US/07857224B
; Patent No. 5958784
; GENERAL INFORMATION:

APPLICANT: Benner, Steven A.
TITLE OF INVENTION: Predicting Folded Structures of Proteins
NUMBER OF SEQUENCES: 114
CORRESPONDENCE ADDRESS:
ADDRESSEE: Steven A. Benner
STREET: Hadlaubstrasse 151
CITY: Zurich
STATE: none
COUNTRY: Switzerland
ZIP: (note: this is an international post code) CH-8092
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch diskette, 1.4 Mb storage
COMPUTER: Apple Macintosh
OPERATING SYSTEM: Macintosh 7.0
SOFTWARE: Microsoft Word
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/857,224B
FILING DATE: 03/25/92
CLASSIFICATION: 436
PRIOR APPLICATION DATA: none
TELECOMMUNICATION INFORMATION:
TELEPHONE: (International) 41 1 632 2830
TELEFAX: (International) 41 1 262 2437
TELEX: none
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 264
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Drosophila melanogaster
FEATURE: Protein kinase; Table 8 Column 18
PUBLICATION INFORMATION:
AUTHORS:
AUTHORS: Hanks, S. K.
AUTHORS: Quinn, A. M.
AUTHORS: Hunter, T.
TITLE: The protein kinase family
JOURNAL: Science
VOLUME: 241
PAGES: 42-52
DATE: 1988
US-07-857-224B-17

Query Match 0.4%; Score 7; DB 2; Length 264;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 227 GKSVDW 233
DB 180 GKSVDW 186

RESULT 28
5290690-10
PATENT NO. 5290690
APPLICANT: MRABET, NADIR; LASTERS, IGNADE; SPANSENS, PATRICK
MATTHYSSENS, GASTON; WODAK, SHOSHANA; QUAX, WILHELMUS J.
TITLE OF INVENTION: METHODS AND MEANS FOR CONTROLLING THE
STABILITY OF PROTEINS
NUMBER OF SEQUENCES: 22
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/398,706
FILING DATE: 25-AUG-1989
SEQ ID NO: 10
LENGTH: 334
5290690-10

Query Match 0.4%; Score 7; DB 6; Length 334;

Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1345 NLVWNGK 1351
DB 63 NLVWNGK 69
RESULT 29
5290690-9
PATENT NO. 5290690
APPLICANT: MRABET, NADIR; LASTERS, IGNADE; SPANSENS, PATRICK
MATTHYSSENS, GASTON; WODAK, SHOSHANA; QUAX, WILHELMUS J.
TITLE OF INVENTION: METHODS AND MEANS FOR CONTROLLING THE
STABILITY OF PROTEINS
NUMBER OF SEQUENCES: 22
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/398,706
FILING DATE: 25-AUG-1989
SEQ ID NO: 9
LENGTH: 335
5290690-9

Query Match 0.4%; Score 7; DB 6; Length 335;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1345 NLVWNGK 1351
DB 64 NLVWNGK 70

RESULT 30
US-08-975-762-70
SEQUENCE 70, Application US/08975762
PATENT NO. 6207169
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Lodes, Michael J.
APPLICANT: Houghton, Raymond
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND
NUMBER OF SEQUENCES: 73
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/975,762
FILING DATE: 21-MAR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: MEKI, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.439
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-622-4900
TELEFAX: 206-682-6031
INFORMATION FOR SEQ ID NO: 70:
SEQUENCE CHARACTERISTICS:
LENGTH: 378 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein

US-08-975-762-70

Query Match 0.4%; Score 7; DB 4; Length 378;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 549 SAVSGFD 555
Db 248 SAVSGFD 254

Search completed: August 12, 2002, 08:10:16
Job time: 167 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 12, 2002, 08:07:59 ; Search time 36.49 Seconds
(without alignments)
4689.918 Million cell updates/sec

Title: US-09-995-749a-2

Perfect score: 1781

Sequence: 1 MEIKKFKLKSQKQWTA.....SDAEYFTSDVGKMLDQNK 1781

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283138 seqs, 96089334 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database :

1: PIR71:*
2: PIR1:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	22	1.2	1431	2 A45866	dextranucrase (EC
2	17	1.0	1599	2 S22737	glucosyltransferas
3	16	0.9	1577	2 T30858	glucosyltransferas
4	15	0.8	1290	2 JC5473	dextranucrase (EC
5	15	0.8	1375	2 JT0345	dextranucrase (EC
6	15	0.8	1475	2 B33135	glfB protein precu
7	15	0.8	1592	2 A38175	glucosyltransferas
8	13	0.7	1365	2 A41483	glucosyltransferas
9	13	0.7	1449	2 T30857	glucosyltransferas
10	13	0.7	1449	2 T30552	glucosyltransferas
11	13	0.7	1508	2 T31098	probable dextranuc
12	13	0.7	1518	2 A44811	glucosyltransferas
13	8	0.4	207	2 AB1877	general secretion
14	8	0.4	228	2 F70934	probable lpgN prot
15	8	0.4	251	2 AH2344	ATP-binding protei
16	8	0.4	274	2 S76999	methionyl aminopep
17	8	0.4	292	2 T09784	homeobox leucine z
18	8	0.4	297	2 G81565	hypothetical prote
19	8	0.4	297	2 E72102	5'-methylthioadeno
20	8	0.4	297	2 H86519	hypothetical prote
21	8	0.4	304	2 A87258	hypothetical prote
22	8	0.4	329	2 A55221	dextranase inhibito
23	8	0.4	335	1 DEBSGF	glyceralddehyde-3-p
24	8	0.4	348	2 S76794	hypothetical prote
25	8	0.4	355	2 AE1279	membrane proteins
26	8	0.4	355	2 AE1642	membrane protein h
27	8	0.4	388	2 T21061	hypothetical prote
28	8	0.4	405	2 D97563	hypothetical prote
29	8	0.4	405	2 AD2784	molybdopterin bios

30	8	0.4	413	2 AB3174	aminotransferase,
31	8	0.4	458	1 C70058	conserved hypothet
32	8	0.4	489	2 T41446	conserved hypothet
33	8	0.4	556	2 E75049	phenylalanine--tRN
34	8	0.4	561	2 C86420	unknown protein, 1
35	8	0.4	629	2 A30168	homeotic protein L
36	8	0.4	635	2 S01164	homeotic protein L
37	8	0.4	682	1 KIBOGC	protein kinase C (
38	8	0.4	697	1 KIRTC	protein kinase C (
39	8	0.4	697	1 KIRBGC	protein kinase C (
40	8	0.4	697	2 JN0348	protein kinase C (
41	8	0.4	697	2 D24664	protein kinase C (
42	8	0.4	716	2 AG2446	single-stranded-DN
43	8	0.4	777	2 C95059	hypothetical prote
44	8	0.4	777	2 B97928	type I site-specif
45	8	0.4	1091	2 S01998	contactin precurs
46	8	0.4	1136	2 A56559	enhancer-trap-locu
47	8	0.4	1331	2 T04938	hypothetical prote
48	8	0.4	1495	2 A85240	hypothetical prote
49	8	0.4	1495	2 T10649	hypothetical prote
50	8	0.4	2265	2 T26183	hypothetical prote
51	8	0.4	3263	2 R82410	hypothetical prote
52	7	0.4	54	1 LBRP2S	light-harvesting p
53	7	0.4	54	2 A27087	light-harvesting p
54	7	0.4	56	2 PD0016	dextranucrase (EC
55	7	0.4	92	2 B97701	ribosomal protein
56	7	0.4	101	2 A97855	hypothetical prote
57	7	0.4	103	2 A53096	interleukin-8 prec
58	7	0.4	105	2 T29157	hypothetical prote
59	7	0.4	112	2 C86374	protein T23E23.22
60	7	0.4	112	2 AB0887	probable membrane
61	7	0.4	119	2 B97196	hypothetical prote
62	7	0.4	134	2 F84367	hypothetical prote
63	7	0.4	134	2 T45188	hypothetical prote
64	7	0.4	134	2 T10555	hypothetical prote
65	7	0.4	138	2 S31513	ig heavy chain - h
66	7	0.4	140	2 S54084	probable membrane
67	7	0.4	141	2 G69386	hypothetical prote
68	7	0.4	144	2 C97042	hypothetical prote
69	7	0.4	144	2 B72280	hypothetical prote
70	7	0.4	146	1 C34588	DNA-directed RNA p
71	7	0.4	150	2 H97158	hypothetical prote
72	7	0.4	152	2 S75230	hypothetical prote
73	7	0.4	162	2 D82802	hypothetical prote
74	7	0.4	163	2 H84039	hypothetical prote
75	7	0.4	164	2 T08346	hypothetical prote
76	7	0.4	166	2 B72607	hypothetical prote
77	7	0.4	174	2 AC3433	hypothetical prote
78	7	0.4	176	2 A84199	hypothetical prote
79	7	0.4	188	2 E71157	probable GDP-alcob
80	7	0.4	188	2 A81987	probable oxidoredu
81	7	0.4	190	2 B81042	conserved hypothet
82	7	0.4	193	1 S39401	hydrogenase matura
83	7	0.4	194	2 C64065	aminoacyl-tRNA hyd
84	7	0.4	196	2 T23657	hypothetical prote
85	7	0.4	203	2 S67198	hypothetical prote
86	7	0.4	204	2 T08661	anti-silencing pro
87	7	0.4	204	2 AC1600	SOS response regul
88	7	0.4	204	2 AF1237	SOS response regul
89	7	0.4	206	1 F71208	hypothetical prote
90	7	0.4	213	2 F97681	ABC-type transport
91	7	0.4	213	2 AE2906	hypothetical prote
92	7	0.4	216	2 C97843	serine esterase ho
93	7	0.4	221	2 A82331	hypothetical prote
94	7	0.4	222	2 T06444	GTP-binding protei
95	7	0.4	222	2 S68832	hypothetical prote
96	7	0.4	223	2 B49591	hypothetical prote
97	7	0.4	224	2 T21788	hypothetical prote
98	7	0.4	224	2 H72346	basal-body rod mod
99	7	0.4	227	2 G70859	probable secreted
100	7	0.4	230	2 D70658	hypothetical prote

ALIGNMENTS

RESULT 1
A45866
dextranucrase (EC 2.4.1.5) precursor - Streptococcus mutans
C:Species: Streptococcus mutans
C>Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 07-May-1999
C:Accession: A45866
R: Honda, O.; Kato, C.; Kuramitsu, H.K.
J. Gen. Microbiol. 136, 2099-2105, 1990
A:Title: Nucleotide sequence of the Streptococcus mutans gtfD gene encoding the glucosyltransferase
A:Reference number: A45866; MUID:91100958
A:Accession: A45866
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1431 <HON>
A:Cross-references: GB:M29296
C:Superfamily: cpl repeat homology
F:181-201/Domain: cpl repeat homology <CP1>
F:1127-1146/Domain: cpl repeat homology <CP2>
F:1192-1211/Domain: cpl repeat homology <CP3>
F:1257-1276/Domain: cpl repeat homology <CP4>
F:1277-1297/Domain: cpl repeat homology <CP5>
F:1321-1340/Domain: cpl repeat homology <CP8>
F:1341-1361/Domain: cpl repeat homology <CP6>
F:1385-1404/Domain: cpl repeat homology <CP7>
Query Match 1.2%; Score 22; DB 2; Length 1431;
Best Local Similarity 100.0%; Pred. No. 1.5e-13;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 979 FILANDIDNSNPVQAEQLNWL 1000
Db 420 FILANDIDNSNPVQAEQLNWL 441
RESULT 2
S22737
glucosyltransferase (EC 2.4.1.-) S - Streptococcus salivarius
C:Species: Streptococcus salivarius
C>Date: 16-Sep-1992 #sequence_revision 16-Sep-1992 #text_change 21-Jan-2000
C:Accession: S22737; S28810; B44811; S22727
R: Jacques, N.
submitted to the EMBL Data Library, March 1992
A:Reference number: S22726
A:Accession: S22737
A:Molecule type: DNA
A:Residues: 1-1599 <JAC>
A:Cross-references: EMBL:211872; NID:g47530; PIDN:CAA77898.1; PID:g47531
A:Experimental source: ATCC 25975
R: Giffard, P.M.; Simpson, C.L.; Milward, C.P.; Jacques, N.A.
J. Gen. Microbiol. 137, 2577-2593, 1991
A:Title: Molecular characterization of a cluster of at least two glucosyltransferase genes
A:Reference number: A44811; MUID:92148377
A:Accession: S28810
A:Molecule type: DNA
A:Residues: 1-51 <GIF>
A:Cross-references: EMBL:211873
C:Genetics:
A:Gene: gtfK
C:Superfamily: cpl repeat homology
C:Keywords: glycosyltransferase; hexosyltransferase
F:1456-1475/Domain: cpl repeat homology <CPR>
Query Match 1.0%; Score 17; DB 2; Length 1599;
Best Local Similarity 100.0%; Pred. No. 2.9e-08;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1560 YGGAFLDKLQKLYPEIF 1576

Db 1011 YGGAFLDKLQKLYPEIF 1027

RESULT 3

T30858
glucosyltransferase - Streptococcus salivarius
C:Species: Streptococcus salivarius
C>Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
C:Accession: T30858
R: Simpson, C.L.; Giffard, P.M.; Jacques, N.A.
Infect. Immun. 63, 609-621, 1995
A:Title: Streptococcus salivarius ATCC 25975 possesses at least two genes coding for
A:Reference number: Z20909; MUID:95122197
A:Accession: T30858
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1577 <SIM>
A:Cross-references: EMBL:L35928; NID:g662380; PID:g662381; PIDN:AAC41413.1
C:Genetics:
A:Gene: gtfm

Query Match 0.9%; Score 16; DB 2; Length 1577;
Best Local Similarity 100.0%; Pred. No. 3.2e-07;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1508 ADMVPDQIYNLPQKEV 1523

Db 1035 ADMVPDQIYNLPQKEV 1050

RESULT 4

JC5473
dextranucrase (EC 2.4.1.5) - Leuconostoc mesenteroides
C:Species: Leuconostoc mesenteroides
C>Date: 07-Jul-1997 #sequence_revision 29-Aug-1997 #text_change 29-Aug-1997
C:Accession: JC5473
R: Monchois, V.; Willemot, R.M.; Remaud-Simeon, M.; Croux, C.; Monsan, P.
Gene 182, 23-32, 1996
A:Title: Cloning and sequencing of a gene coding for a novel dextranucrase from Leuc
A:Reference number: JC5473; MUID:97136886
A:Accession: JC5473
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1230 <MON>
A:Cross-references: GB:U38181
C:Comment: This enzyme catalyzes the transfer of D-glucopyranosyl units from sucrose
C:Genetics:
A:Gene: dsrA
C:Keywords: glycosyltransferase; hexosyltransferase
F:78-870/Domain: catalytic #status predicted <CAT>
F:922-1290/Domain: glucan-binding #status predicted <GCB>

Query Match 0.8%; Score 15; DB 2; Length 1290;
Best Local Similarity 100.0%; Pred. No. 3e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 986 DNSNPVQAEQLNWL 1000

Db 248 DNSNPVQAEQLNWL 262

RESULT 5

JT0345
dextranucrase (EC 2.4.1.5) precursor - Streptococcus mutans (strain GS-5)
N:Alternate names: sucrose 6-glucosyltransferase
C:Species: Streptococcus mutans
C>Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 24-Sep-1999
C:Accession: JT0345; C33135
R: Ueda, S.; Shiroza, T.; Kuramitsu, H.K.
Gene 69, 101-109, 1988

A:Title: Sequence analysis of the gtfc gene from Streptococcus mutans GS-5.

A:Reference number: JT0345; MUID:89137980
 A:Accession: JT0345
 A:Molecule type: DNA
 A:Residues: 1-1375 <UED>
 A:Experimental source: GS-5
 R:Shiroza, T.; Ueda, S.; Kuramitsu, H.K.
 J. Bacteriol. 169, 4263-4270, 1987
 A:Title: Sequence analysis of the gtfc gene from Streptococcus mutans.
 A:Reference number: A33135; MUID:87308013
 A:Accession: C33135
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-349 <SH>
 A:Cross-references: GB:M17361
 C:Genetics:
 A:Gene: gtfc
 C:Function:
 A:Description: catalyzes the synthesis of both water-soluble and water-insoluble glucans
 C:Superfamily: cpl repeat homology
 C:Keywords: duplication; glycosyltransferase; hexosyltransferase
 F:1-34/Domain: signal sequence #status predicted <SIG>
 F:35-1375/Product: glycosyltransferase #status predicted <MAT>
 F:1126-1145/Domain: cpl repeat homology <CP1>
 F:1253-1272/Domain: cpl repeat homology <CP2>
 F:1318-1337/Domain: cpl repeat homology <CP3>

Query Match 0.8%; Score 15; DB 2; Length 1375;

Best Local Similarity 100.0%; Pred. No. 3.2e-06; Mismatches 0; Indels 0; Gaps 0;

Matches 15; Conservative 0;

QY 986 DNSNPVVQAEQLNWL 1000

DB 439 DNSNPVVQAEQLNWL 453

RESULT 6

B33135
 gtfc protein precursor - Streptococcus mutans
 C:Species: Streptococcus mutans
 C:Date: 23-Oct-1990 #sequence_revision 23-Oct-1990 #text_change 15-Oct-1999
 C:Accession: B33135; A33128
 R:Shiroza, T.; Ueda, S.; Kuramitsu, H.K.
 J. Bacteriol. 169, 4263-4270, 1987
 A:Title: Sequence analysis of the gtfc gene from Streptococcus mutans.
 A:Reference number: A33135; MUID:87308013
 A:Accession: B33135
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1475 <SH>
 A:Cross-references: GB:M17361; NID:g153639; PIDN:AAA88588.1; PID:g153640
 R:Shiroza, T.; Ueda, S.; Kuramitsu, H.K.
 submitted to the Protein Sequence Database, September 1990
 A:Reference number: A33128
 A:Accession: A33128
 A:Status: preliminary; not compared with conceptual translation
 A:Molecule type: DNA
 A:Residues: 1-171,173-641,N',643-1475 <SH2>
 A:Experimental source: strain GS-5
 C:Superfamily: cpl repeat homology
 F:1096-1115/Domain: cpl repeat homology <CP1>
 F:1224-1243/Domain: cpl repeat homology <CP2>
 F:1289-1308/Domain: cpl repeat homology <CP3>
 F:1354-1373/Domain: cpl repeat homology <CP4>
 F:1419-1438/Domain: cpl repeat homology <CP5>

Query Match 0.8%; Score 15; DB 2; Length 1475;

Best Local Similarity 100.0%; Pred. No. 3.4e-06; Mismatches 0; Indels 0; Gaps 0;

Matches 15; Conservative 0;

QY 986 DNSNPVVQAEQLNWL 1000

DB 413 DNSNPVVQAEQLNWL 427

RESULT 7

A38175
 glucosyltransferase precursor - Streptococcus sobrinus
 C:Species: Streptococcus sobrinus
 C:Date: 28-Aug-1992 #sequence_revision 28-Aug-1992 #text_change 15-Oct-1999
 C:Accession: A38175
 R:Abo, H.; Msumura, T.; Kodama, T.; Ohta, H.; Fukui, K.; Kato, K.; Kagawa, H.
 J. Bacteriol. 173, 989-996, 1991
 A:Title: Peptide sequences for sucrose splitting and glucan binding within Streptococ
 A:Reference number: A38175; MUID:91123227
 A:Accession: A38175
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1592 <ABO>
 A:Cross-references: GB:D90213; NID:g217032; PIDN:BAA14241.1; PID:d1014946; PID:g21703

C:Superfamily: cpl repeat homology
 F:1093-1112/Domain: cpl repeat homology <CP1>
 F:1222-1241/Domain: cpl repeat homology <CP2>
 F:1287-1306/Domain: cpl repeat homology <CP3>
 F:1330-1351/Domain: cpl repeat homology <CP4>
 F:1352-1371/Domain: cpl repeat homology <CP5>
 F:1402-1420/Domain: cpl repeat homology <CP6>
 F:1465-1484/Domain: cpl repeat homology <CP7>
 F:1513-1532/Domain: cpl repeat homology <CP8>

Query Match 0.8%; Score 15; DB 2; Length 1592;

Best Local Similarity 100.0%; Pred. No. 3.6e-06; Mismatches 0; Indels 0; Gaps 0;

Matches 15; Conservative 0;

QY 986 DNSNPVVQAEQLNWL 1000

DB 409 DNSNPVVQAEQLNWL 423

RESULT 8

A41483
 glucosyltransferase (EC 2.4.1.-) gtfs precursor - Streptococcus sobrinus
 C:Species: Streptococcus sobrinus
 C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 15-Oct-1999
 C:Accession: A41483
 R:Gimre, K.S.; Russell, R.R.B.; Ferretti, J.J.
 Infect. Immun. 58, 2452-2458, 1990
 A:Title: Analysis of the Streptococcus downei gtfs gene, which specifies a glucosyltr
 A:Reference number: A41483; MUID:90316665
 A:Accession: A41483
 A:Molecule type: DNA
 A:Residues: 1-1365 <GII>
 A:Cross-references: GB:M30943; NID:g153652; PIDN:AAA26898.1; PID:g153653
 C:Genetics:
 A:Gene: gtfs
 C:Superfamily: cpl repeat homology
 C:Keywords: glycosyltransferase; hexosyltransferase

Query Match 0.7%; Score 13; DB 2; Length 1365;

Best Local Similarity 100.0%; Pred. No. 0.0004; Mismatches 0; Indels 0; Gaps 0;

Matches 13; Conservative 0;

QY 986 DNSNPVVQAEQLN 998

DB 399 DNSNPVVQAEQLN 411

RESULT 9

T30857
 glucosyltransferase - Streptococcus salivarius
 C:Species: Streptococcus salivarius
 C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999

C:Accession: T30857
R:Simpson, C.L.; Giffard, P.M.; Jacques, N.A.
Infect. Immun. 63, 609-621, 1995
A:Title: Streptococcus salivarius ATCC 25975 possesses at least two genes coding for pil
A:Reference number: 220909; MUID:95122197
A:Accession: T30857
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1449 <SIN>
A:Cross-references: EMBL:L35495; NID:g662378; PID:g662379; PIDN:AAC41412.1
C:Genetics:
A:Gene: gtfL

Query Match 0.7%; Score 13; DB 2; Length 1449;
Best Local Similarity 100.0%; Pred. No. 0.00042;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 978 EFLANDIDNSNP 990
|||||
Db 460 EFLANDIDNSNP 472

RESULT 10
T30552
glucosyltransferase N - Streptococcus salivarius (fragment)
C:Species: Streptococcus salivarius
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
C:Accession: T30552
K:Jaife, R.I.
submitted to the EMBL Data Library, February 1998
A:Description: Streptococcus salivarius V1477 gtfN.
A:Reference number: 220854
A:Accession: T30552
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1449 <JAF>
A:Cross-references: EMBL:AF049609; NID:g2935545; PID:g2935546; PIDN:AAC05156.1
C:Genetics:
A:Gene: gtfN

Query Match 0.7%; Score 13; DB 2; Length 1449;
Best Local Similarity 100.0%; Pred. No. 0.00042;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 978 EFLANDIDNSNP 990
|||||
Db 460 EFLANDIDNSNP 472

RESULT 11
T31098
Probable dextranucrase (EC 2.4.1.5), extracellular - Leuconostoc mesenteroides
C:Species: Leuconostoc mesenteroides
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 11-May-2000
C:Accession: T31098
R:Monchois, V.; Renaud-Simeon, M.; Monsan, P.; Willemot, R.M.
FEMS Microbiol. Lett. 159, 307-315, 1998
A:Title: Cloning and sequencing of a gene coding for an extracellular dextranucrase (DS
A:Reference number: 220981; MUID:98164374
A:Accession: T31098
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1508 <MON>
A:Cross-references: EMBL:AF030129; NID:g2766611; PID:g2766612; PIDN:AAB95453.1
A:Experimental source: strain NRRL B-1299
C:Genetics:
A:Gene: dsrB
C:Function:
A:Description: produces dextran composed only of alpha(1-6) glucosidic bonds
C:Keywords: glycosyltransferase; hexosyltransferase

Query Match 0.7%; Score 13; DB 2; Length 1508;
Best Local Similarity 100.0%; Pred. No. 0.00043;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1508 ADWVPQIYNLPG 1520
|||||
Db 1004 ADWVPQIYNLPG 1016

RESULT 12
A44811
glucosyltransferase (EC 2.4.1.-) I - Streptococcus salivarius
C:Species: Streptococcus salivarius
C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 15-Oct-1999
C:Accession: A44811; S22726; S28809
R:Giffard, P.M.; Simpson, C.L.; Milward, C.P.; Jacques, N.A.
J. Gen. Microbiol. 137, 2577-2593, 1991
A:Title: Molecular characterization of a cluster of at least two glucosyltransferase
A:Reference number: A44811; MUID:92148377
A:Accession: A44811
A:Molecule type: DNA
A:Residues: 1-1518 <GIF>
A:Cross-references: EMBL:Z11873; NID:g47526; PIDN:CAA77900.1; PID:g47527
A:Note: sequence extracted from NCBI backbone (NCBIN:81050, NCBIP:81052)
C:Genetics:
A:Gene: gtfJ
C:Superfamily: cpl repeat homology
C:Keywords: glycosyltransferase; hexosyltransferase
F:1307-1326/Domain: cpl repeat homology <CP4>

Query Match 0.7%; Score 13; DB 2; Length 1518;
Best Local Similarity 100.0%; Pred. No. 0.00043;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1371 AVWVPVGASDND 1383
|||||
Db 858 AVWVPVGASDND 870

RESULT 13
AB1877
general secretion pathway protein G [imported] - Anabaena sp. (strain PCC 7120)
C:Species: Anabaena sp.
A:Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 11-Jan-2002
C:Accession: AB1877
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Irigui
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AB1877
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-207 <KUR>
A:Cross-references: GB:BA000019; PIDN:BAB72521.1; PID:g17129908; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: all0563

Query Match 0.4%; Score 8; DB 2; Length 207;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 386 ATDLSELQ 393
|||||
Db 111 ATDLSELQ 118

RESULT 14

F70934
Probable lpgN protein - Mycobacterium tuberculosis (strain H37Rv)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: F70934
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holtroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295967
A:Accession: F70934
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-228 <COL>
A:Cross-references: GB:AL021942; GB:AL123456; NID:g3242298; PIDN:CAAI7454.1; PID:el25309
A:Experimental source: strain H37Rv
C:Genetics:
A:Gene: lpgN

Query Match 0.4%; Score 8; DB 2; Length 228;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 TAAVATVA 25
IIIIIIII
Db 5 TAAVATVA 12

RESULT 15
AH2344
ATP-binding protein of ABC transporter alr4311 [imported] - Anabaena sp. (strain PCC 7120)
C:Species: Anabaena sp.
A:Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 11-Jan-2002
C:Accession: AH2344
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S. DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AH2344
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-251 <KOR>
A:Cross-references: GB:BA000019; PIDN:BA876010.1; PID:g17133447; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr4311

Query Match 0.4%; Score 8; DB 2; Length 251;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 462 NOTASNID 489
IIIIIIII
Db 224 NOTASNID 231

RESULT 16
S76999
methionyl aminopeptidase (EC 3.4.11.18) 2 - Synecocystis sp. (strain PCC 6803)
N:Alternate names: protein slr0786
C:Species: Synecocystis sp.
A:Variety: PCC 6803
C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 01-Feb-2002
C:Accession: S76999
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, M.; O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, M. DNA Res. 3, 109-136, 1996

A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synecocystis s.
A:Reference number: S74322; MUID:97061201
A:Accession: S76999
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-274 <KAN>
A:Cross-references: EMBL:D64005; GB:AB001339; NID:g1001779; PIDN:BAAI0691.1; PID:g100
A:Note: The nucleotide sequence was submitted to the EMBL Data Library, June 1996
C:Genetics:
A:Gene: map2
A:Start codon: GTG
C:Superfamily: Escherichia coli methionyl aminopeptidase
C:Keywords: aminopeptidase; cobalt; metalloprotein; protein biosynthesis
F:120,131,258/Binding site: cobalt 2 (Asp, Asp, Glu) #status predicted

Query Match 0.4%; Score 8; DB 2; Length 274;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 825 PKQILKDG 832
IIIIIIII
Db 107 PKQILKDG 114

RESULT 17
T09784
homeobox leucine zipper protein Hb-2, dehydration-inducible - Craterostigma plantagin
C:Species: Craterostigma plantagineum
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 09-Jun-2000
C:Accession: T09784
R:Frank, W.; Phillips, J.; Salamini, F.; Bartels, D.
Plant J. 15, 413-421, 1999
A:Title: Two dehydration-inducible transcripts from the resurrection plant Craterostigma plantagin
A:Reference number: Z16852
A:Accession: T09784
A:Status: translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-292 <FRA>
A:Cross-references: EMBL:AJ005833; NID:el294850; PID:el294851
C:Genetics:
A:Gene: hb-2
C:Superfamily: unassigned homeobox proteins; homeobox homology
C:Keywords: DNA binding; homeobox; nucleus; stress-induced protein; transcription reg

Query Match 0.4%; Score 8; DB 2; Length 292;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 309 VGAANS GF 316
IIIIIIII
Db 266 VGAANS GF 273

RESULT 18
G81565
hypothetical protein CP0532 [imported] - Chlamydomophila pneumoniae (strain AR39)
C:Species: Chlamydomophila pneumoniae, Chlamydia pneumoniae
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 11-May-2000
C:Accession: G81565
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke, C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg, Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis Mofn and Chlamydia pneumoniae AR39
A:Reference number: AB1500; MUID:20150255
A:Accession: G81565
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-297 <REA>
A:Cross-references: GB:AE002213; GB:AE002161; NID:g7189446; PIDN:AAF38356.1; PID:g718
A:Experimental source: strain AR39, HL cells

C:Genetics:
A:Gene: CP0532

Query Match 0.4%; Score 8; DB 2; Length 297;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1568 LQKLYPEI 1575
DB 223 LQKLYPEI 230

RESULT 19

E72102 5'-methylthioadenosine nucleosidase homolog - Chlamydothila pneumoniae (strain CWL029)

C:Species: Chlamydothila pneumoniae, Chlamydia pneumoniae
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 05-May-2000
C:Accession: E72102
R:Kallman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.; Nature Genet. 21, 385-389, 1999
A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
A:Reference number: A72000; MUID:99206606
A:Accession: E72102
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-297 <ARN>
A:Cross-references: GB:AE001609; GB:AE001363; NID:g4376502; PIDN:AA018385.1; PTD:g437650

C:Experimental source: strain CWL029
C:Genetics:
A:Gene: CP0232

Query Match 0.4%; Score 8; DB 2; Length 297;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1568 LQKLYPEI 1575
DB 223 LQKLYPEI 230

RESULT 20

H86519 hypothetical protein CPJ0232 [imported] - Chlamydothila pneumoniae (strain J138)

C:Species: Chlamydothila pneumoniae, Chlamydia pneumoniae
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 02-Mar-2001
C:Accession: H86519
R:Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; Is Nucleic Acids Res. 28, 2311-2314, 2000
A:Title: Comparison of whole genome sequences of Chlamydia pneumoniae J138.
A:Reference number: A86491; MUID:20330349
A:Accession: H86519
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-297 <STO>
A:Cross-references: GB:BA000008; NID:g8978605; PIDN:BA098442.1; GSPDB:GN00142
C:Genetics:
A:Gene: CPJ0232

Query Match 0.4%; Score 8; DB 2; Length 297;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1568 LQKLYPEI 1575
DB 223 LQKLYPEI 230

RESULT 21

A87258 hypothetical protein CC0074 [imported] - Caulobacter crescentus

C:Species: Caulobacter crescentus

C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C:Accession: A87258
R:Neuman, W.C.; Feldblyum, T.V.; Paulsen, I.R.; Nelson, K.E.; Eisen, J.; Heidelberg, B.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Ko n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete genome sequence of Caulobacter crescentus.
A:Reference number: A87249; MUID:21173698; PMID:11259647

A:Accession: A87258

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-304 <STO>

A:Cross-references: GB:AE005673; NID:gl3421171; PIDN:AAK22061.1; GSPDB:GN00148

C:Genetics:

A:Gene: CC0074

Query Match 0.4%; Score 8; DB 2; Length 304;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 879 NADADPAE 886
DB 8 NADADPAE 15

RESULT 22

A55221

dextranase inhibitor precursor - Streptococcus sobrinus (strain UA866, serotype g)

C:Species: Streptococcus sobrinus

C:Date: 23-Mar-1995 #sequence_revision 23-Mar-1995 #text_change 17-Mar-1999

C:Accession: A55221

R:Sun, J.W.; Wanda, S.Y.; Camilli, A.; Curtiss III, R.

J. Bacteriol. 176, 7213-7222, 1994

A:Title: Cloning and DNA sequencing of the dextranase inhibitor gene (dei) from Strep

A:Reference number: A55221; MUID:95050304

A:Accession: A55221

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-329 <SUN>

A:Cross-references: GB:L34406

C:Superfamily: cpl repeat homology

F:173-192/Domain: cpl repeat homology <CP1>

Query Match 0.4%; Score 8; DB 2; Length 329;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 737 GLQTINGQ 744
DB 185 GLQTINGQ 192

RESULT 23

DB559F

glyceraldehyde-3-phosphate dehydrogenase (EC 1.2.1.12) [validated] - Bacillus stearot

N:Alternate names: triosephosphate dehydrogenase

C:Species: Bacillus stearotherophilus

C:Date: 24-Apr-1984 #sequence_revision 17-Nov-2000 #text_change 17-Nov-2000

C:Accession: J50164; PS0343; A93186; A91096; A00374

R:Branlant, C.; Oster, T.; Branlant, G.

Gene 75, 145-155, 1989

A:Title: Nucleotide sequence determination of the DNA region coding for Bacillus stea

n in Escherichia coli.

A:Reference number: J50164; MUID:89252911

A:Accession: J50164

A:Molecule type: DNA

A:Residues: 1-335 <BRA>

A:Cross-references: GB:M24493; NID:gl42951; PIDN:AAA22461.1; PTD:gl42952

R:Davies, G.J.; Littlechild, J.A.; Watson, H.C.; Hall, L.

Gene 109, 39-45, 1991

```

RESULT 25
AE1279
membrane proteins homolog lmo1637 [imported] - Listeria monocytogenes (strain EGD-e)
A:Reference number: JQ1399; MUID:92097950
A:Accession: PS0343
A:Molecule type: DNA
A:Residues: 315-335 <DAV>
A:Cross-references: EMBL:X58059; NID:g48853; PIDN:CRAA1092.1; PID:g48854
A:Experimental source: strain NCAL1503
R:Biesecker, G.; Harris, J.I.; Thiery, J.C.; Walker, J.E.; Wonacott, A.J.
Nature 266, 328-333, 1977
A:Title: Sequence and structure of D-glyceraldehyde 3-phosphate dehydrogenase from Bacil
A:Reference number: A93186; MUID:7717126
A:Contents: sequence; X-ray crystallography, 2.7 angstroms
A:Accession: A93186
A:Molecule type: protein
A:Residues: 2-36, 'N', 38, 'DG', 41-58, 'V', 60-61, 'DGDVS', 67-78, 'N', 80-124, 'V', 126, 'N', 128-12
R:Walker, J.E.; Carne, A.F.; Runswick, M.J.; Bridgen, J.; Harris, J.I.
Eur. J. Biochem. 108, 549-565, 1980
A:Title: D-Glyceraldehyde-3-phosphate dehydrogenase. Complete amino-acid sequence of the
A:Reference number: A91096; MUID:81003878
A:Accession: A91096
A:Molecule type: protein
A:Residues: 2-36, 'N', 38, 'DG', 41-58, 'V', 60-61, 'DGDVS', 67-78, 'N', 80-124, 'V', 126, 'N', 128-12
C:Genetics:
A:Gene: gap
C:Superfamily: glyceraldehyde-3-phosphate dehydrogenase
C:Keywords: gluconeogenesis; glycolysis; homotetramer; NAD; oxidoreductase
F:4-34/Region: beta-alpha-beta NAD nucleotide-binding fold
F:152/Active site: Cys #status experimental
F:179/Active site: His #status predicted

Query Match 0.4%; Score 8; DB 1; Length 335;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1345 NLVNGKE 1352
|||||||
Db 64 NLVNGKE 71

RESULT 24
S76794
hypothetical protein - Synecocystis sp. (strain PCC 6803)
C:Species: Synecocystis sp.
A:Variety: PCC 6803
C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 26-May-2000
C:Accession: S76794
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
O.K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synecocystis
s.
A:Reference number: S74322; MUID:97061201
A:Accession: S76794
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-348 <KAN>
A:Cross-references: EMBL:D90916; GB:AB001339; NID:g1653715; PIDN:BAAL8706.1; PID:d101943
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C:Superfamily: unassigned WD repeat proteins; WD repeat homology
F:268-301/Domain: WD repeat homology <WD1>
F:310-343/Domain: WD repeat homology <WD2>

Query Match 0.4%; Score 8; DB 2; Length 348;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 678 SGFQGIIT 685
|||||||
Db 58 SGFQGIIT 65

RESULT 26
AE1642
membrane protein homolog lin1678 [imported] - Listeria innocua (strain Clip11262)
C:Species: Listeria innocua
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
C:Accession: AE1642
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Blo
.; Dominguez-Bernal, G.; Buchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Kreft, J.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehl
ok, C.; Schlueter, T.; Simeos, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehl
A:Title: Comparative genomics of Listeria species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AE1279
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-355 <GLA>
A:Cross-references: GB:NC_003210; PIDN:CAC99715.1; PID:g16411073; GSPDB:GN00177
A:Experimental source: strain EGD-e
C:Genetics:
A:Gene: lmo1637

Query Match 0.4%; Score 8; DB 2; Length 355;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1436 ADFFASLG 1443
|||||||
Db 142 ADFFASLG 149

RESULT 27
T21061
hypothetical protein F17C11.9 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
C:Accession: T21061
R:McMurray, A.

```

```

RESULT 25
AE1279
membrane proteins homolog lmo1637 [imported] - Listeria monocytogenes (strain EGD-e)
A:Reference number: JQ1399; MUID:92097950
A:Accession: PS0343
A:Molecule type: DNA
A:Residues: 315-335 <DAV>
A:Cross-references: EMBL:X58059; NID:g48853; PIDN:CRAA1092.1; PID:g48854
A:Experimental source: strain NCAL1503
R:Biesecker, G.; Harris, J.I.; Thiery, J.C.; Walker, J.E.; Wonacott, A.J.
Nature 266, 328-333, 1977
A:Title: Sequence and structure of D-glyceraldehyde 3-phosphate dehydrogenase from Bacil
A:Reference number: A93186; MUID:7717126
A:Contents: sequence; X-ray crystallography, 2.7 angstroms
A:Accession: A93186
A:Molecule type: protein
A:Residues: 2-36, 'N', 38, 'DG', 41-58, 'V', 60-61, 'DGDVS', 67-78, 'N', 80-124, 'V', 126, 'N', 128-12
R:Walker, J.E.; Carne, A.F.; Runswick, M.J.; Bridgen, J.; Harris, J.I.
Eur. J. Biochem. 108, 549-565, 1980
A:Title: D-Glyceraldehyde-3-phosphate dehydrogenase. Complete amino-acid sequence of the
A:Reference number: A91096; MUID:81003878
A:Accession: A91096
A:Molecule type: protein
A:Residues: 2-36, 'N', 38, 'DG', 41-58, 'V', 60-61, 'DGDVS', 67-78, 'N', 80-124, 'V', 126, 'N', 128-12
C:Genetics:
A:Gene: lmo1637

Query Match 0.4%; Score 8; DB 2; Length 355;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1436 ADFFASLG 1443
|||||||
Db 142 ADFFASLG 149

RESULT 26
AE1642
membrane protein homolog lin1678 [imported] - Listeria innocua (strain Clip11262)
C:Species: Listeria innocua
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
C:Accession: AE1642
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Blo
.; Dominguez-Bernal, G.; Buchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Kreft, J.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehl
ok, C.; Schlueter, T.; Simeos, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehl
A:Title: Comparative genomics of Listeria species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AE1642
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-355 <GLA>
A:Cross-references: GB:AL592022; PIDN:CAC96909.1; PID:g16414165; GSPDB:GN00178
A:Experimental source: strain Clip11262
C:Genetics:
A:Gene: lin1678

Query Match 0.4%; Score 8; DB 2; Length 355;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1436 ADFFASLG 1443
|||||||
Db 142 ADFFASLG 149

RESULT 27
T21061
hypothetical protein F17C11.9 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
C:Accession: T21061
R:McMurray, A.

```

submitted to the EMBL Data Library, May 1996

A:Reference number: Z19366

A:Accession: T21061

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-398 <WILL>

A:Cross-references: EMBL:Z72507; PIDN:CAA96631.1; GSPDB:GN00023; CESP:F17C11.9

A:Experimental source: clone F17C11

C:Genetics:

A:Gene: CESP:F17C11.9

A:Map position: 5

A:introns: 3/3; 44/3; 151/3; 196/1; 354/3

C:Superfamily: translation elongation factor eEF-1 gamma chain

Query Match 0.4%; Score 8; DB 2; Length 398;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 568 LMGQLQVL 575

DB 113 LMGQLQVL 120

RESULT 28

D97563 hypothetical protein AGR_C3105 [imported] - Agrobacterium tumefaciens (strain C58, Cere

C:Species: Agrobacterium tumefaciens

C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 11-Jan-2002

C:Accession: D97563

R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Ouello, B.; Goldman,

A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Leppas, C.; Markelz, B.;

Science 294, 2323-2328, 2001

A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum

A:Reference number: A97359; PMID:11743194

A:Accession: D97563

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-405 <KUR>

A:Cross-references: GB:AE007869; PIDN:AAK87461.1; PID:g15156781; GSPDB:GN00169

C:Genetics:

A:Gene: AGR_C3105

A:Map position: circular chromosome

C:Superfamily: molybdenum cofactor biosynthesis protein moeA-2

Query Match 0.4%; Score 8; DB 2; Length 405;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 457 RFSADENG 464

DB 351 RFSADENG 358

RESULT 29

AD2784 molybdopter in biosynthesis protein [imported] - Agrobacterium tumefaciens (strain C58, D

C:Species: Agrobacterium tumefaciens

C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 01-Feb-2002

C:Accession: AD2784

R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L

erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClell

; Karp, P.; Romero, P.; Zhang, S.

Science 294, 2317-2323, 2001

A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,

ster, E.W.

A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.

A:Reference number: AB2577; PMID:11743193

A:Accession: AD2784

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-405 <KUR>

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 12, 2002, 08:09:54 ; Search time 18.99 seconds

(Without alignments)

3631.360 Million cell updates/sec

Title: US-09-995-749a-2

Perfect score: 1781

Sequence: 1 MEIKKFKLYKSKQWYTA.....SDAEYPTSDVGKMLDQNKL 1781

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 105224 seqs, 38719550 residues

Word size : 0

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database : SwisProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	22	1.2	1462	1	GTED_STRMU
2	15	0.8	1375	1	GTFC_STRMU
3	15	0.8	1476	1	GTEB_STRMU
4	15	0.8	1592	1	GTEF_STRMU
5	14	0.8	1597	1	GTEI_STRMU
6	13	0.7	1365	1	GTEF_STRMU
7	8	0.4	1499	1	GLPA_PANTR
8	8	0.4	191	1	RR4_CYAME
9	8	0.4	274	1	AMP2_SYNY3
10	8	0.4	334	1	G3P_BACST
11	8	0.4	348	1	YE91_SYNY3
12	8	0.4	398	1	EF1G_CAEEL
13	8	0.4	556	1	SYFB_PYRAS
14	8	0.4	635	1	HMLA_DROME
15	8	0.4	682	1	KPGC_BOVIN
16	8	0.4	697	1	KPGC_HUMAN
17	8	0.4	697	1	KPGC_HUMAN
18	8	0.4	697	1	KPGC_RABIT
19	8	0.4	725	1	ADBE_MOUSE
20	8	0.4	1010	1	CONT_CHICK
21	8	0.4	1068	1	H1PR_MOUSE
22	7	0.4	54	1	LHA2_RHOSH
23	7	0.4	91	1	RS19_BACSU
24	7	0.4	103	1	IL8_PIG
25	7	0.4	124	1	IB15_HALNI
26	7	0.4	129	1	RK19_PROWI
27	7	0.4	129	1	YQ09_HALNI
28	7	0.4	141	1	YA96_ARCFU
29	7	0.4	146	1	RPB8_YEAST
30	7	0.4	161	1	R1SB_STRCO
31	7	0.4	188	1	Y460_PYRHO
32	7	0.4	192	1	A16_ANOGA
33	7	0.4	193	1	HUPD_BRAJA

RESULT 1

ALIGNMENTS

34	7	0.4	194	1	PTH_HAEIN
35	7	0.4	203	1	Y080_YEAST
36	7	0.4	237	1	Y194_AQARE
37	7	0.4	243	1	TONB_ENTAE
38	7	0.4	243	1	TONB_KLEPN
39	7	0.4	266	1	NUCM_ERWCH
40	7	0.4	288	1	Y587_PASMU
41	7	0.4	289	1	AMIA_ECOLI
42	7	0.4	289	1	AMIA_SALTY
43	7	0.4	299	1	HEY1_MOUSE
44	7	0.4	299	1	YWFK_BACSU
45	7	0.4	300	1	RIMK_ECOLI
46	7	0.4	300	1	RIMK_SALTY
47	7	0.4	304	1	HEY1_CANPA
48	7	0.4	304	1	HEY1_HUMAN
49	7	0.4	312	1	RT02_ACACA
50	7	0.4	319	1	PA1_SERLI
51	7	0.4	322	1	Y320_BP1F1
52	7	0.4	323	1	DBHX_MOUSE
53	7	0.4	334	1	G3P1_BACSU
54	7	0.4	334	1	G3P1_BACME
55	7	0.4	334	1	G3P_CIOAB
56	7	0.4	334	1	G3P_CLOPA
57	7	0.4	338	1	G3P_TRIHA
58	7	0.4	338	1	REG2_YEAST
59	7	0.4	339	1	G3P_HAEIN
60	7	0.4	350	1	YC73_METJA
61	7	0.4	363	1	ECT1_SCHPO
62	7	0.4	365	1	PAC1_SCHPO
63	7	0.4	376	1	G4P1_YEAST
64	7	0.4	380	1	KAPB_YEAST
65	7	0.4	390	1	YHBZ_ECOLI
66	7	0.4	396	1	YG78_METJA
67	7	0.4	398	1	KAPC_YEAST
68	7	0.4	407	1	YKR5_CAEEL
69	7	0.4	413	1	GLYA_BACHD
70	7	0.4	428	1	MANA_YEAST
71	7	0.4	445	1	OST4_CAEEL
72	7	0.4	453	1	KYNU_YEAST
73	7	0.4	463	1	NRTR_MOUSE
74	7	0.4	463	1	ROK_HUMAN
75	7	0.4	463	1	ROK_RABIT
76	7	0.4	464	1	NRTR_HUMAN
77	7	0.4	464	1	ROK_MOUSE
78	7	0.4	465	1	NRTR_CHICK
79	7	0.4	466	1	HEMN_SYNY3
80	7	0.4	472	1	YMF2_BACSU
81	7	0.4	482	1	RBMT_ARATH
82	7	0.4	486	1	VN53_ROTTH
83	7	0.4	486	1	VN53_ROTTH
84	7	0.4	517	1	VU1_HPV12
85	7	0.4	518	1	GTTL_YEAST
86	7	0.4	518	1	LEUL_BACSU
87	7	0.4	536	1	ZF94_RAT
88	7	0.4	548	1	PPAC_THEME
89	7	0.4	572	1	MOES_LYTVA
90	7	0.4	584	1	SACB_ACEOI
91	7	0.4	592	1	INV2_DAUCA
92	7	0.4	608	1	YD6C_SCHPO
93	7	0.4	639	1	KPCL_DROME
94	7	0.4	643	1	DXS_MYCLE
95	7	0.4	670	1	C21D_HUMAN
96	7	0.4	671	1	KPCL_HUMAN
97	7	0.4	671	1	KPCL_RABIT
98	7	0.4	671	1	KPCL_RAT
99	7	0.4	672	1	KPCA_BOVIN
100	7	0.4	672	1	KPCA_HUMAN

P44682	haemophilus
Q08746	saccharomyc
Q67734	aquifex aeo
P46383	enterobacte
P45610	klebsiella
P37994	erwinia chr
Q9cn56	pasteurella
P36548	escherichia
P33772	salmonella
Q9wv93	mus musculus
P39647	bacillus su
P17116	escherichia
Q9c521	salmonella
Q9tsz2	canis famil
Q9y5j3	homo sapien
P46753	acanthamoeb
P18952	serriata li
Q80302	bacteriopho
P70694	mus musculus
P09124	bacillus su
P23722	bacillus me
O52631	clostridium
Q89309	clostridium
P87197	trichoderma
P38232	saccharomyc
P44304	haemophilus
Q58669	methanococc
P22192	schizosacch
Q9uti6	schizosacch
P46672	saccharomyc
P06245	saccharomyc
P42641	escherichia
Q59072	methanococc
P05986	saccharomyc
P34311	caenorhabdi
Q9k694	bacillus ha
P29952	saccharomyc
P45971	caenorhabdi
Q05979	saccharomyc
O08842	mus musculus
Q07244	homo sapien
O19049	oryctolagus
O00451	homo sapien
O60577	mus musculus
O13157	gallus gall
P74132	synecocyst
P39641	bacillus su
Q9x184	a probable
P30212	human rotav
P35424	human papil
P36733	human papil
P25346	saccharomyc
P94565	bacillus su
Q9z2k3	rattus norv
Q9wz56	thermotoga
P52962	lytechinus
Q43998	acetobacter
Q39692	daucus caro
P10324	schizosacch
P05130	drosophila
Q50000	mycobacteri
Q95447	homo sapien
P05771	homo sapien
P05772	oryctolagus
P04410	rattus norv
P04409	bos taurus
P17252	homo sapien

GTFFD_STRMU STANDARD; PRT: 1462 AA.
AC G49331; 069386; 069389; 069392; 069398;
DT 01-FEB-1996 (Rel. 33, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Glucosyltransferase-S precursor (EC 2.4.1.5) (GTF-S) (Dextranucrase)
DE (Sucrose 6-glucosyltransferase).
GN GTFD.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GS-5;
RX MEDLINE=91100958; PubMed=2148600;
RA Honda O., Kato C., Kuramitsu H.K.;
RT "Nucleotide sequence of the Streptococcus mutans gtfD gene encoding
the glucosyltransferase-S enzyme.";
RL J. Gen. Microbiol. 136:2099-2105(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MT4239, MT4245, MT4251, MT4467, AND MT8148;
RX MEDLINE=98231643; PubMed=9570124;
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Sobue S.,
Kumai S., Hamada S.;
RT "Molecular analyses of glucosyltransferase genes among strains of
Streptococcus mutans.";
RL FEMS Microbiol. Lett. 161:331-336(1998).
CC -!- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT
TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE
OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE
AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.
CC -!- CATALYTIC ACTIVITY: Sucrose + {(1,6)-alpha-D-glucosyl}(N) = D-
fructose + {(1,6)-alpha-D-glucosyl}(N+1).
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- DISEASE: DENTAL CARIES.
CC -!- MISCELLANEOUS: GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA
1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES
WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH
FORMS OF GLUCANS.
CC -!- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-
BINDING PROTEIN FROM S. MUTANS.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@sib-sib.ch).

DR EMBL; M29296; AAA26895.1; -
DR EMBL; D88653; BAA26103.1; -
DR EMBL; D88656; BAA26107.1; -
DR EMBL; D88659; BAA26111.1; -
DR EMBL; D88662; BAA26115.1; -
DR EMBL; D89979; BAA26121.1; -
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 1.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Transferase; Glycosyltransferase; Signal; Repeat; Dental caries.
FT SIGNAL 1 ? POTENTIAL.
FT CHAIN ? GLUCOSYLTRANSFERASE-S.
FT DOMAIN 1232 1423 3 x 63 AA APPROXIMATE TANDEM REPEATS.
FT REPEAT 1232 1295 1.
FT REPEAT 1296 1359 2.
FT REPEAT 1360 1423 3.
FT VARIANT 58 68 K -> E (IN STRAIN MT4467).
FT VARIANT 68 68 A -> S (IN STRAINS MT4239 AND MT4245).
FT VARIANT 81 81 A -> T (IN STRAINS MT4251 AND MT8148).

FT VARIANT 113 113 T -> I (IN STRAINS MT4239 AND MT4245).
FT VARIANT 122 122 A -> V (IN STRAINS MT4239, MT4245 AND
MT8148).
FT VARIANT 132 132 S -> A (IN STRAINS MT4239, MT4245, MT4251
AND MT8148).
FT VARIANT 135 135 A -> V (IN STRAIN MT4245).
FT VARIANT 202 202 V -> L (IN STRAIN MT4239).
FT VARIANT 255 255 D -> N (IN STRAIN MT8148).
FT VARIANT 275 275 E -> D (IN STRAINS MT4239, MT4245 AND
MT4251).
FT VARIANT 288 288 D -> N (IN STRAINS MT4239, MT4245 AND
MT4251).
FT VARIANT 301 301 Q -> H (IN STRAIN MT4245).
FT VARIANT 313 313 D -> N (IN STRAINS MT4239 AND MT4251).
FT VARIANT 317 317 E -> K (IN STRAIN MT4239).
FT VARIANT 328 328 V -> F (IN STRAIN MT4239).
FT VARIANT 350 350 F -> L (IN STRAINS MT4239, MT4251 AND
MT4467).
FT VARIANT 628 628 KKQYTO -> EKEYTL (IN STRAIN MT4251).
FT VARIANT 688 688 A -> S (IN STRAIN MT4239).
FT VARIANT 726 726 TDQSEA -> ADKGNDS (IN STRAIN MT4251).
FT VARIANT 730 TDQGS -> ADKGN (IN STRAINS MT4239 AND
MT4245).
FT VARIANT 964 964 D -> Y (IN STRAIN MT4251).
FT VARIANT 1019 1019 E -> K (IN STRAINS MT4245 AND MT4251).
FT VARIANT 1059 1060 LG -> IR (IN STRAIN MT4251).
FT VARIANT 1060 1060 G -> R (IN STRAIN MT4245).
FT VARIANT 1080 1080 Q -> H (IN STRAINS MT4239, MT4245,
MT4251, MT4467 AND MT8148).
FT VARIANT 1142 1142 S -> N (IN STRAIN MT4239).
FT VARIANT 1198 1198 Y -> C (IN STRAINS MT4251 AND MT4467).
FT VARIANT 1220 1220 F -> L (IN STRAIN MT4467).
FT VARIANT 1280 1280 Q -> P (IN STRAIN MT4245).
FT VARIANT 1282 1282 K -> T (IN STRAIN MT4245).
FT VARIANT 1290 1290 N -> D (IN STRAIN MT4245).
FT VARIANT 1311 1311 D -> G (IN STRAINS MT4239, MT4245, MT4251
AND MT8148).
FT VARIANT 1403 1403 R -> G (IN STRAINS MT4239, MT4245,
MT4251, MT4467 AND MT8148).
FT VARIANT 1425 1425 R -> K (IN STRAIN MT4467).
FT VARIANT 1449 1449 RYDKNSGMVYKVVTLANGRRIGIDRWGIARY ->
FT CONFLICT 1428 1462 VYR (IN REF. 1).
FT SEQUENCE 1462 AA; 163512 MW; 5C5541F0DC80DF00 CRC64;
Query Match 1.28; Score 22; DB 1; Length 1462;
Best Local Similarity 100.0%; Pred. No. 7.1e-14;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 979 FLANDIDNSNPVQAEQLNWL 1000
Db 420 FLANDIDNSNPVQAEQLNWL 441
RESULT 2
GTFD_STRMU STANDARD; PRT: 1375 AA.
AC GTFD_STRMU STANDARD; PRT: 1375 AA.
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Glucosyltransferase-SI precursor (EC 2.4.1.5) (GTF-SI)
DE (Dextranucrase) (Sucrose 6-glucosyltransferase).
GN GTFD.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GS-5;
RX MEDLINE=89137980; PubMed=2976010;


```
FT VARIANT 89 89 S -> F (IN STRAIN MT4251).
FT VARIANT 168 168 K -> N (IN STRAIN MT4251).
FT VARIANT 276 276 D -> S (IN STRAINS MT4239, MT4245 AND
MT4251).
FT VARIANT 399 399 N -> R (IN STRAIN MT4239).
FT VARIANT 474 474 I -> T (IN STRAIN MT4239).
FT VARIANT 512 512 K -> R (IN STRAIN MT4239).
FT VARIANT 519 519 F -> Y (IN STRAIN MT4239).
FT VARIANT 701 701 T -> Y (IN STRAIN MT4239).
FT VARIANT 708 708 A -> V (IN STRAIN MT4239).
FT VARIANT 938 938 F -> L (IN STRAIN MT4239).
FT VARIANT 952 957 YGTPVA -> FGKPE (IN STRAINS MT4245,
MT4251 AND MT4248).
FT VARIANT 963 964 NT -> SV (IN STRAINS MT4245, MT4251 AND
MT4248).
FT VARIANT 968 970 VDG -> ADS (IN STRAINS MT4245, MT4251 AND
MT4248).
FT VARIANT 1086 1086 A -> T (IN STRAIN MT4239).
FT VARIANT 1158 1158 S -> N (IN STRAIN MT4239).
FT VARIANT 1163 1163 H -> Y (IN STRAIN MT4251).
FT VARIANT 1168 1168 E -> K (IN STRAIN MT4251).
FT VARIANT 1182 1182 Y -> C (IN STRAIN MT4251).
FT VARIANT 1234 1234 A -> P (IN STRAIN MT4239).
FT VARIANT 1263 1263 H -> P (IN STRAIN MT4239).
FT VARIANT 1263 1263 H -> R (IN STRAINS MT4239, MT4245 AND
MT4251).
FT VARIANT 1264 1264 H -> Y (IN STRAINS MT4245 AND MT4251).
FT VARIANT 1272 1272 G -> S (IN STRAINS MT4245 AND MT4251).
FT VARIANT 1329 1329 Y -> S (IN STRAINS MT4239, MT4245, MT4251
AND MT4248).
FT VARIANT 1394 1394 H -> Y (IN STRAINS MT4245 AND MT4251).
FT VARIANT 1402 1402 G -> S (IN STRAINS MT4245 AND MT4251).
FT VARIANT 1459 1459 Y -> H (IN STRAIN MT4467).
FT CONFLICT 570 570 R -> A (IN REF. 1).
FT CONFLICT 800 817 ADQDVRAAPSTPDGK -> LKMFALRLPHQOMA
(IN REF. 1).
FT CONFLICT 1310 1310 H -> L (IN REF. 1).
FT CONFLICT 1476 1476 H -> L (IN REF. 1).
SQ SEQUENCE 1476 AA; 165685 MW; 3479B62B07694D98 CRC64;

Query Match 0.8%; Score 15; DB 1; Length 1476;
Best Local Similarity 100.0%; Pred. No. 1.4e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 986 DNSNPFVQAEQLNWL 1000
Db 413 DNSNPFVQAEQLNWL 427

RESULT 4
GTF2_STRDO
ID GTF2_STRDO STANDARD; PRT; 1592 AA.
AC P27470.
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Glucosyltransferase-I precursor (EC 2.4.1.5) (GTF-I) (Dextranase)
DE (Sucrose 6-glucosyltransferase).
OS Streptococcus downei (Streptococcus sobrinus).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1317;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=6715;
RX MEDLINE=91123227; PubMed=1704006;
RA Abo H., Matsumura T., Kodama T., Ohta H., Fukui K., Kato K.,
RA Kagawa H.;
RT "Peptide sequences for sucrose splitting and glucan binding within
RT Streptococcus sobrinus glucosyltransferase (water-insoluble glucan
RT synthetase).";
RL J. Bacteriol. 173:989-996(1991).
CC --J- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT
```

```
CC TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE
CC OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE
CC AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.
CC -I- CATALYTIC ACTIVITY: Sucrose + ((1,6)-alpha-D-glucosyl)(N) - D-
CC fructose + ((1,6)-alpha-D-glucosyl)(N+1).
CC -I- SUBCELLULAR LOCATION: Secreted.
CC -I- DISEASE: DENTAL CARIES.
CC -I- MISCELLANEOUS: GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA
CC 1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES
CC WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH
CC FORMS OF GLUCANS.
CC -I- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-
CC BINDING PROTEIN FROM S. MUTANS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D90213; BAA14241.1; --
CC PIR; A38175; A38175.
CC HSSP; P00695; 2HEE
CC InterPro; IPR002479; CW_binding.
CC InterPro; IPR003318; Glyco_hydro_70.
CC Pfam; PF01473; CW_binding_1; 16.
CC Pfam; PF02324; Glyco_hydro_70; 1.
CC Transferase; Glycosyltransferase; Signal; Repeat; Dental caries.
KW SIGNAL
FT CHAIN 1 38 POTENTIAL.
FT CHAIN 39 1592 GLUCOSYLTRANSFERASE-I.
FT DOMAIN 39 1044 CATALYTIC (APPROXIMATE).
FT DOMAIN 1093 1592 GLUCAN-BINDING (APPROXIMATE).
FT REPEAT 1093 1592 6.5 X TANDEM REPEATS.
FT REPEAT 1158 1207 1.
FT REPEAT 1222 1272 3.
FT REPEAT 1287 1337 4.
FT REPEAT 1402 1451 5.
FT REPEAT 1514 1563 6.
FT REPEAT 1577 1592 7 (INCOMPLETE).
SQ SEQUENCE 1592 AA; 176167 MW; BCOA66D079351ECF CRC64;

Query Match 0.8%; Score 15; DB 1; Length 1592;
Best Local Similarity 100.0%; Pred. No. 1.5e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 986 DNSNPFVQAEQLNWL 1000
Db 409 DNSNPFVQAEQLNWL 423

RESULT 5
GTF1_STRDO
ID GTF1_STRDO STANDARD; PRT; 1597 AA.
AC P11001.
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Glucosyltransferase-I precursor (EC 2.4.1.5) (GTF-I) (Dextranase)
DE (Sucrose 6-glucosyltransferase).
OS Streptococcus downei (Streptococcus sobrinus).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1317;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MFE28;
RX MEDLINE=87308014; PubMed=3040686;
RA Ferretti J.J., Gilpin M.L., Russell R.R.B.;
```

RT "Nucleotide sequence of a glucosyltransferase gene from Streptococcus
 RL sobrinus MFe28";
 CC J. Bacteriol. 169:4271-4278(1987).
 CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT
 CC TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE
 CC OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE
 CC AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.
 CC -1- CATALYTIC ACTIVITY: Sucrose + ((1,6)-alpha-D-glucosyl)(N) - D-
 CC fructose + ((1,6)-alpha-D-glucosyl)(N+1).
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- DISEASE: DENTAL CARIES.
 CC -1- MISCELLANEOUS: GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA
 CC 1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES
 CC WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH
 CC FORMS OF GLUCANS.
 CC -1- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-
 CC BINDING PROTEIN FROM S. MUTANS.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch)
 CC -----
 DR EMBL: M17391; AAC63063.1; -;
 DR InterPro: IPR002479; CW_Binding.
 DR InterPro: IPR003318; Glyco_hydro_70.
 DR Pfam: PF01473; CW_Binding_1; 19.
 DR Pfam: PF02324; Glyco_hydro_70; 1.
 DR Transferase; Glycosyltransferase; Signal; Repeat; Dental caries.
 FT SIGNAL 1 38 POTENTIAL.
 FT CHAIN 39 1597 GLUCOSYLTRANSFERASE-I.
 FT DOMAIN 39 1050 CATALYTIC (APPROXIMATE).
 FT DOMAIN 1099 1597 GLUCAN-BINDING (APPROXIMATE).
 FT DOMAIN 1099 1597 1.25 A, 2 B, AND 5 AC REPEATS.
 FT REPEAT 1099 1132 A REPEAT.
 FT REPEAT 1163 1213 AC REPEAT.
 FT REPEAT 1227 1277 AC REPEAT.
 FT REPEAT 1292 1342 AC REPEAT.
 FT REPEAT 1352 1399 B REPEAT.
 FT REPEAT 1406 1455 AC REPEAT.
 FT REPEAT 1465 1512 B REPEAT.
 FT REPEAT 1519 1568 AC REPEAT.
 FT REPEAT 1582 1597 A REPEAT (INCOMPLETE).
 SQ SEQUENCE 1597 AA; 177080 MW; B9E86A200868798E CRC64;

Query Match 0.8%; Score 14; DB 1; Length 1597;
 Best Local Similarity 100.0%; Pred. No. 1.7e-05;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1468 NGVAFTRDYDLGMS 1481
 |||||
 Db 891 NGVAFTRDYDLGMS 904

RESULT 6
 GTF5_STRDO
 ID GTF5_STRDO STANDARD; PRT; 1365 AA.
 AC P29336;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DT 01-APR-1993 (Rel. 25, Last annotation update)
 DE Glycosyltransferase-S precursor (EC 2.4.1.5) (Gtf-S) (Dextranucrase)
 DE (Sucrose 6-glucosyltransferase).
 GN GTF5.
 OS Streptococcus downei (Streptococcus sobrinus).
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 CC Streptococcus.
 OX NCBI_TaxID=1317;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN-MFE28;
 RX MEDLINE=90316665; PubMed=2142479;
 RA Gilmore K.S., Russell R.R., Ferretti J.J.;
 RT "Analysis of the Streptococcus downei gtfis gene, which specifies a
 RT glucosyltransferase that synthesizes soluble glucans";
 RL Infect. Immun. 58:2452-2458(1990).
 CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT
 CC TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE
 CC OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE
 CC AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.
 CC -1- CATALYTIC ACTIVITY: Sucrose + ((1,6)-alpha-D-glucosyl)(N) - D-
 CC fructose + ((1,6)-alpha-D-glucosyl)(N+1).
 CC -1- ENZYME REGULATION: GLUCAN SYNTHESIS BY GTF-S IS INDEPENDENT OF
 CC PRIMER GLUCAN UNLIKE GTF-I.
 CC -1- DISEASE: DENTAL CARIES.
 CC -1- MISCELLANEOUS: GTF-S SYNTHESIZES WATER-SOLUBLE GLUCANS (ALPHA
 CC 1,6-GLUCOSE).
 CC -1- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-
 CC BINDING PROTEIN FROM S. MUTANS.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch)
 CC -----
 DR EMBL: M30943; AAA26898.1; -;
 DR PIR: A41483; A41483.
 DR InterPro: IPR002479; CW_Binding.
 DR InterPro: IPR003318; Glyco_hydro_70.
 DR Pfam: PF01473; CW_Binding_1; 10.
 DR Pfam: PF02324; Glyco_hydro_70; 1.
 DR Transferase; Glycosyltransferase; Signal; Repeat; Dental caries.
 FT SIGNAL 1 36 OR 37 (POTENTIAL).
 FT CHAIN 37 1385 GLUCOSYLTRANSFERASE-S.
 FT DOMAIN 37 1050 CATALYTIC (APPROXIMATE).
 FT DOMAIN 1083 1365 GLUCAN-BINDING (APPROXIMATE).
 FT DOMAIN 1083 1365 4.5 X TANDEM REPEATS.
 FT REPEAT 1083 1131 1.
 FT REPEAT 1150 1199 2.
 FT REPEAT 1225 1274 3.
 FT REPEAT 1289 1339 4.
 FT REPEAT 1353 1365 5 (INCOMPLETE).
 SQ SEQUENCE 1365 AA; 151590 MW; 167296B5A2E8C476 CRC64;

Query Match 0.7%; Score 13; DB 1; Length 1365;
 Best Local Similarity 100.0%; Pred. No. 0.00016;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 986 DNSNPVVQAEQLN 998
 |||||
 Db 399 DNSNPVVQAEQLN 411

RESULT 7
 GLPA_PANTR
 ID GLPA_PANTR STANDARD; PRT; 149 AA.
 AC Q28913;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Glycophorin A precursor.
 DE GYPA OR GPA.
 GN Pan troglodytes (Chimpanzee).
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
 OX NCBI_TaxID=9598;
 RN [1]
 RP SEQUENCE FROM N.A.

```
RX MEDLINE-96029146; PubMed-7563135;
RA Huang C.H., Xie S.S., Socha W., Blumenfeld O.O.;
RT "Sequence diversification and exon inactivation in the glycoprotein A
RL gene family from chimpanzee to human.";
RJ J. Mol. Evol. 41:478-486(1995).
CC -!- FUNCTION: GLYCOPHORIN A IS THE MAJOR INTRINSIC MEMBRANE PROTEIN OF
CC THE ERYTHROCYTE.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- SIMILARITY: BELONGS TO THE GLYCOPHORIN A FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: S79724; AAB35338.1; -.
CC HSSP: P02724; 1MSR.
CC InterPro: IPR001195; Glycophorin_A.
CC Pfam: PF01102; Glycophorin_A; 1.
CC PROSITE: PS00312; GLYCOPHORIN_A; 1.
CC Erythrocyte; Transmembrane; Sialic acid; Glycoprotein; Signal.
CC SIGNAL 1 19
CC CHAIN 20 149
CC DOMAIN 20 90
CC TRANSMEM 91 113
CC DOMAIN 114 149
CC CYTOPLASMIC (BY SIMILARITY).
CC CARBOHYD 21 21
CC O-LINKED (GALNAC. .) (BY SIMILARITY).
CC CARBOHYD 22 22
CC O-LINKED (GALNAC. .) (BY SIMILARITY).
CC CARBOHYD 23 23
CC O-LINKED (GALNAC. .) (BY SIMILARITY).
CC CARBOHYD 29 29
CC O-LINKED (GALNAC. .) (BY SIMILARITY).
CC CARBOHYD 30 30
CC O-LINKED (GALNAC. .) (BY SIMILARITY).
CC CARBOHYD 31 31
CC O-LINKED (GALNAC. .) (BY SIMILARITY).
CC CARBOHYD 32 32
CC O-LINKED (GALNAC. .) (BY SIMILARITY).
CC CARBOHYD 35 35
CC O-LINKED (GALNAC. .) (BY SIMILARITY).
CC CARBOHYD 37 37
CC O-LINKED (GALNAC. .) (BY SIMILARITY).
CC CARBOHYD 40 40
CC O-LINKED (GALNAC. .) (BY SIMILARITY).
CC CARBOHYD 43 43
CC O-LINKED (GALNAC. .) (BY SIMILARITY).
CC CARBOHYD 44 44
CC O-LINKED (GALNAC. .) (BY SIMILARITY).
CC CARBOHYD 51 51
CC O-LINKED (GALNAC. .) (BY SIMILARITY).
CC CARBOHYD 55 55
CC O-LINKED (GALNAC. .) (BY SIMILARITY).
CC CARBOHYD 62 62
CC O-LINKED (GALNAC. .) (BY SIMILARITY).
CC CARBOHYD 68 68
CC O-LINKED (GALNAC. .) (BY SIMILARITY).
CC SEQUENCE 149 AA; 16403 MW; 9AB52EC62FA34607 CRC64;
Query Match 0.4%; Score 8; DB 1; Length 149;
Best Local Similarity 100.0%; Pred. No. 3.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 595 VQLVHRFS 702
DB 80 VQLVHRFS 87
RESULT 8
RR4_CYAME STANDARD; PRI: 191 AA.
AC 022020;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Chloroplast 30S ribosomal protein S4.
GN RPS4.
OS Cyanidioschyzon merolae.
OG Chloroplast.
OC Eukaryota; Rhodophyta; Bangiophyceae; Cyanidioschyzon.
OX NCBI_TaxID=45157;
RN [1]
RP SEQUENCE FROM N.A.
RA Ohta N.;
```

```
RT "Analysis of a plastid gene cluster reveals a close relationship
between Cyanidioschyzon and Cyanidium.";
RL J. Plant Res. 110:235-245(1997).
CC -!- FUNCTION: THIS PROTEIN BINDS DIRECTLY TO 16S RIBOSOMAL RNA
CC (BY SIMILARITY).
CC -!- SIMILARITY: CONTAINS 1 S4 RNA-BINDING DOMAIN.
CC -!- SIMILARITY: BELONGS TO THE S4P FAMILY OF RIBOSOMAL PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D63675; BAA22816.1; -.
CC HSSP: P81288; 1C05.
CC InterPro: IPR001912; Ribosomal_S4.
CC Pfam: PF00163; Ribosomal_S4; 1.
CC Pfam: PF01479; S4; 1.
CC SMART; SM00363; S4; 1.
CC PROSITE; PS00632; RIBOSOMAL_S4; 1.
CC Ribosomal protein; rRNA-binding; Chloroplast.
CC DOMAIN 92 139
CC RNA-BINDING (SA TYPE).
CC SEQUENCE 191 AA; 21758 MW; AUCDF4BB15DF5E61 CRC64;
Query Match 0.4%; Score 8; DB 1; Length 191;
Best Local Similarity 100.0%; Pred. No. 4.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 610 VKVNGNQV 617
DB 119 VKVNGNQV 126
RESULT 9
AMP2_SYNY3 STANDARD; PRI: 274 AA.
ID AMP2_SYNY3
AC P53580;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative methionine aminopeptidase B (EC 3.4.11.18) (MAP) (Peptidase
DE M).
GN SLR0786.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1146;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-96127529; PubMed-8590279;
RA Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,
RA Sugita M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
RT region from map positions 64% to 92% of the genome.";
RL DNA Res. 2:153-166(1995).
CC -!- FUNCTION: REMOVES THE AMINO-TERMINAL METHIONINE FROM NASCENT
CC PROTEINS (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: L-METHIONYLPEPTIDE + H(2)O = L-METHIONINE +
CC PEPTIDE.
CC -!- COFACTOR: COBALT; BINDS 2 IONS PER SUBUNIT, SODIUM; BINDS 1 ION
CC PER SUBUNIT (By similarity).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M24a; ALSO KNOWN AS THE
CC MAP FAMILY 1.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
```

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL: D64005; BAA10691.1; -
 DR HSSP: P07906; IC24.
 DR MEROPS: M24.001; -
 DR InterPro: IPR002467; MAP_1.
 DR InterPro: IPR001714; Methamino_Pfase.
 DR InterPro: IPR000994; Peptidase_M24.
 DR Pfam: PF00557; Peptidase_M24; 1.
 DR PRINTS: PR00599; MAPEPTIDASE.
 DR PROSITE: PS00680; MAP_1; 1.
 DR Hypothetical protein; Hydrolase; Aminopeptidase; Cobalt;
 KW Complete proteome.
 FT METAL 120 120 COBALT 2 (BY SIMILARITY).
 FT METAL 131 131 COBALT 1 AND 2 (BY SIMILARITY).
 FT METAL 194 194 COBALT 1 (BY SIMILARITY).
 FT METAL 227 227 COBALT 1 (BY SIMILARITY).
 FT METAL 258 258 COBALT 1 AND 2 (BY SIMILARITY).
 SQ SEQUENCE 274 AA; 29634 MW; 5DE1A5745A2CB8E CRC64;
 Query Match 0.4%; Score 8; DB 1; Length 274;
 Best Local Similarity 100.0%; Pred. No. 6.5;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 825 PKQILKDG 832
 D 107 PKQILKDG 114
 RESULT 10
 G3P_BACST STANDARD; PRT; 334 AA.
 AC P00362;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Glyceraldehyde 3-phosphate dehydrogenase (EC 1.2.1.12) (GAPDH).
 GN GAP.
 OS Bacillus stearothermophilus.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Geobacillus.
 OX NCBI_TaxID=1422;
 RN [1]
 RP SEQUENCE FROM N.A.
 RY MEDLINE=90060813; PubMed=2684782;
 RA Tesfay H.S., Amelunxen R.E., Goldberg I.D.;
 RT "Nucleotide sequences of genes encoding heat-stable and heat-labile
 RT glyceraldehyde-3-phosphate dehydrogenases; amino acid sequence and
 RT protein thermostability.;
 RL Gene 82:237-248(1989).
 RN [2]
 RP ERRATUM.
 RY MEDLINE=91033059; PubMed=2227448;
 RA Tesfay H.S., Amelunxen R.E., Goldberg I.D.;
 RL Gene 94:144-144(1990).
 RN [3]
 RP SEQUENCE FROM N.A.
 RY MEDLINE=89252911; PubMed=2656407;
 RA Brantant C., Oster T., Brantant G.;
 RT "Nucleotide sequence determination of the DNA region coding for
 RT Bacillus stearothermophilus glyceraldehyde-3-phosphate dehydrogenase
 RT and of the flanking DNA regions required for its expression in
 RT Escherichia coli.;
 RL Gene 75:145-155(1989).
 RN [4]
 RP SEQUENCE.
 RY MEDLINE=81003878; PubMed=7408868;
 RA Walker J.E., Carne A.F., Ruswick M.J., Bridgen J., Harris J.I.;
 RT "D-glyceraldehyde-3-phosphate dehydrogenase. Complete amino-acid
 RT sequence of the enzyme from Bacillus stearothermophilus.";

RL Eur. J. Biochem. 108:549-565(1980).
 RN [5]
 RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS).
 RX MEDLINE=77171226; PubMed=193030;
 RA Biesecker G., Harris J.I., Thierry J.C., Walker J.E., Wonacott A.J.;
 RT "Sequence and structure of D-glyceraldehyde 3-phosphate dehydrogenase
 RT from Bacillus stearothermophilus.";
 RL Nature 266:328-333(1977).
 RN [6]
 RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).
 RX MEDLINE=87226185; PubMed=3586018;
 RA Skarzynski T., Moody P.C.E., Wonacott A.J.;
 RT "Structure of holo-glyceraldehyde-3-phosphate dehydrogenase from
 RT Bacillus stearothermophilus at 1.8-A resolution.";
 RL J. Mol. Biol. 193:171-187(1987).
 RN [7]
 RP X-RAY CRYSTALLOGRAPHY (2.45 ANGSTROMS).
 RX MEDLINE=97318921; PubMed=9175858;
 RA Didierjean C., Rahuel-Clermont S., Vitoux B., Dideberg O.,
 RA Brantant G., Aubry A.;
 RT "A crystallographic comparison between mutated glyceraldehyde-3-
 RT phosphate dehydrogenases from Bacillus stearothermophilus complexed
 RT with either NAD+ or NADP+.";
 RL J. Mol. Biol. 268:739-759(1997).
 CC -1- CATALYTIC ACTIVITY: D-glyceraldehyde 3-phosphate + phosphate +
 CC NAD(+) -> 3-phospho-D-glyceroyl phosphate + NADH.
 CC -1- PATHWAY: FIRST STEP IN THE SECOND PHASE OF GLYCOLYSIS.
 CC -1- SUBUNIT: HOMOTETRAMER.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- SIMILARITY: BELONGS TO THE GLYCERALDEHYDE 3-PHOSPHATE
 CC DEHYDROGENASE FAMILY.
 CC -1- CAUTION: REF.1 SEQUENCE WAS INCORRECT AND RETRACTED IN REF.2.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL: M24493; AAA22461.1; -
 DR PIR: A00374; DEBSGF.
 DR PIR: JS0164; JS0164.
 DR PDB: 1GDI; 15-OCT-89.
 DR PDB: 2GDI; 15-OCT-89.
 DR PDB: 1DBV; 07-JUL-97.
 DR PDB: 2DBV; 07-JUL-97.
 DR PDB: 3DBV; 07-JUL-97.
 DR PDB: 4DBV; 07-JUL-97.
 DR InterPro: IPR000173; GAP_DH.
 DR Pfam: PF00044; gpdh; 1.
 DR Pfam: PF02800; gpdh_C; 1.
 DR PRINTS: PR00078; G3PDHGRNASE.
 DR PROSITE: PS00071; GAPDH; 1.
 KW Glycolysis; Oxidoreductase; NAD; 3D-structure.
 FT INIT MET 0 0
 FT BINDING 151 151 GLYCERALDEHYDE 3-PHOSPHATE.
 FT ACT_SITE 178 178 ACTIVATES THIOL GROUP DURING CATALYSIS.
 FT STRAND 2 7
 FT HELIX 11 22
 FT STRAND 26 32
 FT HELIX 38 46
 FT TURN 49 51
 FT STRAND 58 61
 FT TURN 62 62
 FT STRAND 63 66
 FT TURN 67 68
 FT STRAND 69 74
 FT HELIX 79 82
 FT HELIX 84 87
 FT TURN 88 88
 FT STRAND 91 94

```

FT STRAND 101 101
FT HELIX 102 105
FT TURN 106 106
FT TURN 107 110
FT TURN 111 112
FT STRAND 115 118
FT STRAND 123 123
FT STRAND 127 128
FT TURN 131 133
FT HELIX 135 137
FT TURN 140 142
FT STRAND 145 147
FT HELIX 151 167
FT STRAND 169 179
FT TURN 182 183
FT TURN 196 197
FT HELIX 200 202
FT STRAND 205 207
FT HELIX 211 217
FT TURN 218 218
FT HELIX 220 222
FT TURN 223 225
FT STRAND 226 233
FT STRAND 239 247
FT HELIX 253 265
FT TURN 266 268
FT STRAND 269 269
FT TURN 270 271
FT STRAND 272 275
FT HELIX 281 284
FT TURN 285 286
FT STRAND 291 294
FT HELIX 295 297
FT STRAND 299 301
FT TURN 302 304
FT STRAND 305 312
FT TURN 314 315
FT HELIX 316 331
FT TURN 332 332
SQ SEQUENCE 334 AA; 35944 MW; 7E8D539E25233A2B CRC64;

Query Match 0.4%; Score 8; DB 1; Length 334;
Best Local Similarity 100.0%; Pred. No. 7.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1345 NLVYNGKE 1352
DB 63 NLVYNGKE 70

RESULT 11
YE91_SVNY3
ID YE91_SVNY3 STANDARD; PRT; 348 AA.
AC P74598;
DI 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical WD-repeat protein SL1491.
GN SL1491.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hirose M., Sugita M., Saito S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA Shinozaki S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
-synechocystis sp. strain PCC6803. II. Sequence determination of the
entire genome and assignment of potential protein-coding regions.";
DNA Res. 3:109-136(1996).
-|- SIMILARITY: CONTAINS 6 WD REPEATS (TRP-ASP DOMAINS).
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR EMBL; D90916; BAB18706.1; -
DR InterPro; IPR001880; WD40.
DR Pfam; PF00400; WD40; 6.
DR PRINTS; PR00320; GPROTEINRPT.
DR SMART; SM00320; WD40; 6.
DR PROSITE; PS00678; WD_REPEATS_1; 2.
DR PROSITE; PS00082; WD_REPEATS_2; 4.
DR PROSITE; PS0294; WD_REPEATS_REGION; 1.
KW Hypothetical protein; Repeat; WD repeat; Complete proteome.
FT REPEAT 59 98
FT REPEAT 142 182 WD 1.
FT REPEAT 185 226 WD 2.
FT REPEAT 229 267 WD 3.
FT REPEAT 270 309 WD 4.
FT REPEAT 312 347 WD 5.
FT REPEAT 348 37347 WD 6.
SQ SEQUENCE 348 AA; 37347 MW; D7399CC43E37466F CRC64;

Query Match 0.4%; Score 8; DB 1; Length 348;
Best Local Similarity 100.0%; Pred. No. 8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 678 SGFQGIIT 685
DB 58 SGFQGIIT 65

RESULT 12
EF1G_CAEEL
ID EF1G_CAEEL STANDARD; PRT; 398 AA.
AC P54412;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable elongation factor 1-gamma (EF-1-gamma) (SEF-1B gamma).
GN F17C11.9.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Murray A.;
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
CC -|- FUNCTION: PROBABLY PLAYS A ROLE IN ANCHORING THE COMPLEX TO OTHER
CELLULAR COMPONENTS.
CC -|- SUBUNIT: EF-1 IS COMPOSED OF FOUR SUBUNITS: ALPHA, BETA,
DELTA, AND GAMMA.
CC -|- SIMILARITY: CONTAINS 1 GST-LIKE DOMAIN.
CC -|- SIMILARITY: CONTAINS 1 EF-1-GAMMA DOMAIN.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR EMBL; 272507; CAA96631.1; -

```



```

DR WormPep; F17C11.9; CE05656.
DR InterPro; IPR001662; EFIG.
DR Pfam; PF00647; EFIG_domain; 1.
DR Pfam; PF00043; GST_C; 1.
DR ProDom; PD006217; EFIG; 1.
DR PROSITE; PS0040; EFIG; 1.
KW Hypothetical protein; Elongation factor; Protein biosynthesis.
FT DOMAIN 1 176 GST.
SQ SEQUENCE 398 AA; 44387 MW; 943174C4CABE96E8 CRC64;

Query Match          0.4%; Score 8; DB 1; Length 398;
Best Local Similarity 100.0%; Pred. No. 9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 568 LMGQLQVL 575
DB 113 LMGQLQVL 120

RESULT 13
SIFF_PYRAB
ID SIFF_PYRAB STANDARD; PRT: 556 AA.
AC Q9UXX2;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Phenylalanyl-tRNA synthetase beta chain (P6.1.1.20) (Phenylalanine--
  tRNA ligase beta chain) (PheRS).
GN PHET OR PAR2427.
OS Pyrococcus abyssi.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
OX NCBI_TaxID=29292;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ORSAY;
RA Heilig R.;
RT "Pyrococcus abyssi genome sequence: insights into archaeal chromosome
  structure and evolution.";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: ATP + L-phenylalanine + tRNA(Phe) = AMP +
  diphosphate + L-phenylalanyl-tRNA(Phe).
CC -1- SUBUNIT: Tetramer of two alpha and two beta chains (by
  similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE PHENYLALANYL-TRNA SYNTHETASE BETA
  CHAIN FAMILY. SUBFAMILY 2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ248287; CAB50290.1; -
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
FT DOMAIN 135 138 POLY-ARG.
SQ SEQUENCE 556 AA; 63949 MW; 06134413484E60DB CRC64;

Query Match          0.4%; Score 8; DB 1; Length 556;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 638 VLNVGKEV 645
DB 516 VLNVGKEV 523

```

```

RESULT 14
EMBL_DROME
ID HMLA_DROME STANDARD; PRT: 635 AA.
AC P10105;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Homeotic labial protein (F24) (F90-2).
GN LAB.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OREGON-R;
RA Mlodzik M., Fjose A., Gehring W.J.;
RT "Molecular structure and spatial expression of a homeobox gene from
  the labial region of the Antennapedia-complex.";
RL EMBO J. 7:2569-2578(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC MEDLINE=89252817; PubMed=2566560;
RA Diederich R.J., Merrill V.K.L., Pultz M.A., Kaufman T.C.;
RT "Isolation, structure, and expression of labial, a homeotic gene of
  the Antennapedia complex involved in Drosophila head development.";
RL Genes Dev. 3:399-414(1989).
RN [3]
RP SEQUENCE OF 500-567 FROM N.A.
RC MEDLINE=86253687; PubMed=3014511;
RA Hoey T., Doyle H.J., Harding K., Wedeen C., Levine M.;
RT "Homeo box gene expression in anterior and posterior regions of the
  Drosophila embryo.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:4809-4813(1986).
CC -1- FUNCTION: SEQUENCE-SPECIFIC TRANSCRIPTION FACTOR WHICH IS PART OF
  A DEVELOPMENTAL REGULATORY SYSTEM ON THE ANTERIOR-POSTERIOR AXIS.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
  SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- SIMILARITY: BELONGS TO THE ANTP FAMILY OF HOMEBOX PROTEINS.
CC "LABIAL" SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X13103; CA331495.1; -
DR EMBL; X13104; CAB57787.1; -
DR EMBL; X13105; CAB57787.1; JOINED.
DR EMBL; X13106; CAB57787.1; JOINED.
DR EMBL; M13570; AA28610.1; -
DR PIR; S01164; S01164.
DR PIR; A30168; A30168.
DR FLYBASE; FBgn0002522; lab.
DR HSSP; P14653; 1B72.
DR InterPro; IPR001827; Antennapedia.
DR InterPro; IPR001356; Homeobox.
DR Pfam; PF00046; Homeobox; 1.
DR PRINTS; PR00024; HOMEBOX.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEBOX_1; 1.
DR PROSITE; PS00032; ANTENAPEDIA; FALSE_NEG.
DR PROSITE; PS00071; HOMEBOX_2; 1.
KW Transcription regulation; Homeobox; DNA-binding;
KW Developmental protein; Nuclear protein; Alternative splicing.
FT DOMAIN 11 18 HIS-RICH.

```

```

FT DOMAIN 42 64 GLN-RICH.
FT DOMAIN 96 111 GLN-RICH.
FT DOMAIN 129 134 SER/THR-RICH.
FT DOMAIN 191 196 GLN-RICH.
FT DOMAIN 306 339 GLN/HIS-RICH.
FT DOMAIN 361 370 SER/THR-RICH.
FT DNA_BIND 507 566 HOMEOBOX.
FT VARSPLIC 407 412 MISSING (IN SHORT ISOFORM).
FT CONFLICT 31 31 S -> A (IN REF. 2).
FT CONFLICT 46 46 P -> H (IN REF. 2).
FT CONFLICT 139 139 K -> T (IN REF. 2).
FT CONFLICT 444 444 D -> G (IN REF. 2).
SQ SEQUENCE 635 AA; 68212 MW; CA501599C2019E54 CRC64;

Query Match 0.4%; Score 8; DB 1; Length 635;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 TAATATVA 25
Db 164 TAATATVA 171

RESULT 15
KPCG_BOVIN
ID KPCG_BOVIN STANDARD; PRT; 682 AA.
AC P05128;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DE 15-JUL-1999 (Rel. 38, Last annotation update)
DE Protein kinase C, gamma type (EC 2.7.1.-) (PKC-gamma) (Fragment).
GN PRKCG.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86289426; PubMed=3755548;
RA Coussens L., Parker P.J., Rhee L., Yang-Feng T.L., Chen E.,
RA Waterfield M.D., Francke U., Ullrich A.;
RT "Multiple, distinct forms of bovine and human protein kinase C
RT suggest diversity in cellular signaling pathways.";
RL Science 233:859-866(1986).
RN [2]
RP REVIEW.
RX MEDLINE=88318921; PubMed=3045562;
RA Nishizuka Y.;
RT "The molecular heterogeneity of protein kinase C and its implications
RT for cellular regulation.";
RL Nature 334:661-665(1986).
CC -!- FUNCTION: THIS IS A CALCIUM-ACTIVATED, PHOSPHOLIPID-DEPENDENT,
CC SERINE- AND THREONINE-SPECIFIC ENZYME.
CC -!- FUNCTION: PKC IS ACTIVATED BY DIACYLGLYCEROL WHICH IN TURN
CC PHOSPHORYLATES A RANGE OF CELLULAR PROTEINS. PKC ALSO SERVES AS
CC THE RECEPTOR FOR PHORBOL ESTERS, A CLASS OF TUMOR PROMOTERS.
CC -!- SIMILARITY: CONTAINS 2 ZINC-DEPENDENT PHORBOL-ESTER AND DAG
CC BINDING DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 C2 DOMAIN.
CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC PKC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC JBL; M13976; AAA30704.1; -.

```

```

DR PIR: C24664; KIBOGC.
DR HSP; P05697; ITBN.
DR InterPro: IPR000008; C2.
DR InterPro: IPR002219; DAG_PE-bind.
DR InterPro: IPR000719; Euk_kinase.
DR InterPro: IPR000961; Kinase_C.
DR InterPro: IPR002290; Ser_thr_kinase.
DR Pfam: PF00188; C2; 1.
DR Pfam: PF00130; DAG_PE-bind; 2.
DR Pfam: PF00069; kinase; 1.
DR Pfam: PF00433; kinase_C; 1.
DR SMART: SM00109; C1; 2.
DR SMART: SM00239; C2; 1.
DR SMART: SM00133; S_TK_X; 1.
DR SMART: SM00220; S_TK; 1.
DR PROSITE: PS00499; C2_DOMAIN_1; 1.
DR PROSITE: PS00004; C2_DOMAIN_2; 1.
DR PROSITE: PS00479; DAG_PE_BIND_DOM_1; 2.
DR PROSITE: PS00081; DAG_PE_BIND_DOM_2; 2.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
KW Calcium-binding; Repeat: ATP-binding; Transferase;
KW Serine/threonine-protein kinase; Phorbol-ester binding; Zinc;
KW phosphorylation.
FT NON_TER 1
FT DOMAIN 21 70 PHORBOL-ESTER AND DAG BINDING 1.
FT DOMAIN 86 135 PHORBOL-ESTER AND DAG BINDING 2.
FT DOMAIN 155 245 C2 DOMAIN.
FT DOMAIN 336 599 PROTEIN KINASE.
FT NP_BIND 342 350 ATP (BY SIMILARITY).
FT BINDING 365 365 ATP (BY SIMILARITY).
FT ACT_SITE 465 465 BY SIMILARITY.
FT MOD_RES 633 633 PHOSPHORYLATION (AUTO-) (POTENTIAL).
FT MOD_RES 640 640 PHOSPHORYLATION (AUTO-) (POTENTIAL).
SQ SEQUENCE 682 AA; 77156 MW; 20392D1188C731C CRC64;

Query Match 0.4%; Score 8; DB 1; Length 682;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 227 GKSVDWWS 234
Db 518 GKSVDWWS 525
1111111111

RESULT 16
KPCG_HUMAN
ID KPCG_HUMAN STANDARD; PRT; 697 AA.
AC P05129;
DT 13-AUG-1987 (Rel. 05, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protein kinase C, gamma type (EC 2.7.1.-) (PKC-gamma).
GN PRKCG OR PKCG.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Cui W.C., Yu L., Chu Y.Y., Wang J., Zheng L.H., Zhou G.J., Zhao S.Y.;
RN Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-318 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=86289426; PubMed=3755548;
RA Coussens L., Parker P.J., Rhee L., Yang-Feng T.L., Chen E.,
RA Waterfield M.D., Francke U., Ullrich A.;
RT "Multiple, distinct forms of bovine and human protein kinase C
RT suggest diversity in cellular signaling pathways.";
RL Science 233:859-866(1986).

```

[3] SEQUENCE OF 162-697 FROM N.A.
 RP TISSUE-Hippocampus;
 RC MEDLINE-93387312; PubMed-8375396;
 RX Kochs G., Meyer D., Hug H., Mayne D., Sarre T.F.;
 RT "Activation and substrate specificity of the human protein kinase C
 alpha and zeta isoenzymes.";
 RL Eur. J. Biochem. 216:597-606(1993).
 [4] VARIANTS C-141; Q-415; D-523 AND S-659.
 RP MEDLINE-98213587; PubMed-9545390;
 RX Al-Maghtheh M., Vilhna E.N., Inghearn C.F., Moore T., Bird A.C.,
 RA Bhattacharya S.S.;
 RT "Segregation of a PKCG mutation in two RP11 families.";
 RL Am. J. Hum. Genet. 62:1248-1252(1998).
 [5] SHOWS THAT THE VARIANTS ARE NOT A CAUSE OF RP11.
 RP MEDLINE-99375047; PubMed-1041600;
 RX Dryja T.P., McEvoy J., McGee T.L., Barson E.L.;
 RA "No mutations in the coding region of the PKCG gene in three families
 with retinitis pigmentosa linked to the RP11 locus on chromosome
 19q.";
 RL Am. J. Hum. Genet. 65:926-928(1999).
 CC -!- FUNCTION: THIS IS A CALCIUM-ACTIVATED, PHOSPHOLIPID-DEPENDENT,
 CC SERINE- AND THREONINE-SPECIFIC ENZYME.
 CC -!- FUNCTION: PKC IS ACTIVATED BY DIACYLGLYCEROL WHICH IN TURN
 CC PHOSPHORYLATES A RANGE OF CELLULAR PROTEINS. PKC ALSO SERVES AS
 CC THE RECEPTOR FOR PHORBOL ESTERS, A CLASS OF TUMOR PROMOTERS.
 CC -!- SIMILARITY: CONTAINS 2 ZINC-DEPENDENT PHORBOL-ESTER AND DAG
 CC BINDING DOMAINS.
 CC -!- SIMILARITY: CONTAINS 1 C2 DOMAIN.
 CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 CC -!- PKC SUBFAMILY.
 CC -!- DATABASE: NAME-Mutations of the PKCG gene;
 CC NOTE- retina International's Scientific Newsletter;
 CC WWW="http://www.retina-international.com/sci-news/prkcgmut.htm".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: AF345987; AAK13533.1; .
 DR EMBL: M13977; AAA60102.1; ALT-TERM.
 DR EMBL: Z15114; CAA78820.1; .
 DR PIR: D24664; D24664.
 DR HSSP: P05697; 1TBN.
 DR MIM: 176990; .
 DR InterPro: IPR000008; C2.
 DR InterPro: IPR002219; DAG_PE-bind.
 DR InterPro: IPR000719; Euk_pkinase.
 DR InterPro: IPR000961; Pkinase_C.
 DR InterPro: IPR002290; ser_thr_pkinase.
 DR Pfam: PF00168; C2; 1.
 DR Pfam: PF00130; DAG_PE-bind; 2.
 DR Pfam: PF00069; pkinase; 1.
 DR Pfam: PF00433; pkinase_C; 1.
 DR PRINTS: PR00360; C2DOMAIN.
 DR PRINTS: PR00008; DAGPEDOMAIN.
 DR SMART: SM00109; C1; 2.
 DR SMART: SM00239; C2; 1.
 DR SMART: SM00133; S-TK_X; 1.
 DR SMART: SM00220; S-TK; 1.
 DR PROSITE: PS00499; C2_DOMAIN_1; 1.
 DR PROSITE: PS50004; C2_DOMAIN_2; 1.
 DR PROSITE: PS00479; DAG_PE_BIND_DOM_1; 2.
 DR PROSITE: PS50081; DAG_PE_BIND_DOM_2; 2.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.

KW Calcium-binding; Repeat: ATP-binding; Transferase;
 KW Serine/threonine-protein kinase; Phorbol-ester binding; zinc;
 KW Phosphorylation; Polymorphism.
 FT DOMAIN 36 85 PHORBOL-ESTER AND DAG BINDING 1.
 FT DOMAIN 101 150 PHORBOL-ESTER AND DAG BINDING 2.
 FT DOMAIN 170 260 C2 DOMAIN.
 FT DOMAIN 351 614 PROTEIN KINASE.
 FT NP_BIND 357 365 ATP (BY SIMILARITY).
 FT BINDING 380 380 ATP (BY SIMILARITY).
 FT ACT_SITE 480 480 BY SIMILARITY.
 FT MOD_RES 648 648 PHOSPHORYLATION (AUTO-) (POTENTIAL).
 FT MOD_RES 655 655 PHOSPHORYLATION (AUTO-) (POTENTIAL).
 FT VARIANT 141 141 R -> C.
 FT VARIANT 415 415 /FTID-VAR_008755.
 FT VARIANT 523 523 /FTID-VAR_008756.
 FT VARIANT 659 659 /FTID-VAR_008757.
 FT VARIANT 659 659 R -> S.
 FT SEQUENCE 697 AA; 78447 MW; 3F911B5BEF713C41 CRC64;
 SQ
 Query Match 0.4%; Score 8; DB 1; Length 697;
 Best Local Similarity 100.0%; Pred. No. 15;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 227 GKSDVQWS 234
 |||||
 DB 533 GKSDVQWS 540
 RESULT 17
 KPCG_MOUSE
 ID KPCG_MOUSE STANDARD; PRG: 697 AA.
 AC P05697;
 DT 01-NOV-1988 (Rel. 09, Created)
 DT 01-NOV-1988 (Rel. 09, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Protein Kinase C, gamma type (EC 2.7.1.-) (PKC-gamma).
 GN PRKCG OR PKCG OR PKCG.
 OS Mus musculus (Mouse), and
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090, 10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SPECIES-Rat; TISSUE-Brain;
 RX MEDLINE-88262515; PubMed-3387228;
 RA Ono Y., Fujii T., Igarashi K., Kikkawa U., Ogita K., Nishizuka Y.;
 RT "Nucleotide sequences of cDNAs for alpha and gamma subtypes of rat
 brain protein kinase C.";
 RL Nucleic Acids Res. 16:5199-5200(1988).
 [2]
 RP SEQUENCE FROM N.A.
 RC SPECIES-Rat;
 RX MEDLINE-86272097; PubMed-3755379;
 RA Knopf J.L., Lee M.-H., Sultzman L.A., Kriz R.W., Loomis C.R.,
 RA Hewick R.M., Bell R.M.;
 RT "Cloning and expression of multiple protein kinase C cDNAs";
 RL Cell 46:491-502(1986).
 [3]
 RP SEQUENCE OF 1-56 FROM N.A.
 RC SPECIES-Rat;
 RX MEDLINE-91060619; PubMed-2246272;
 RA Chen K.H., Wilden S.G., Wilson S.H., Huang K.P.;
 RT "Characterization of the 5'-flanking region of the rat protein kinase
 C gamma gene.";
 RL J. Biol. Chem. 265:19961-19965(1990).
 [4]
 RP SEQUENCE FROM N.A.
 RC SPECIES-Mouse; TISSUE-Brain;

RX MEDLINE-93154595; PubMed-8428669;
 RA Bowers B.J., Parham C.L., Sikela J.M., Wehner J.M.;
 RT "Isolation and sequence of a mouse brain cDNA coding for protein
 RL kinase C-gamma isozyme.";
 RN Gene 123:263-265(1993).
 [5]
 RP SEQUENCE FROM N.A.
 RC SPECIES-Mouse; STRAIN-BALB/C; TISSUE-Brain;
 RA Tseng C.P., Verma A.;
 RL Submitted (JAN-1994) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RP STRUCTURE BY NMR OF 91-172.
 RC SPECIES-Rat;
 RX MEDLINE-97419134; PubMed-9271501;
 RA Xu R.X., Pawelczyk T., Xia T.-H., Brown S.C.;
 RT "NMR structure of a protein kinase C-gamma phorbol-binding domain and
 RL study of protein-lipid micelle interactions.";
 RN Biochemistry 36:10709-10717(1997).
 CC -!- FUNCTION: THIS IS A CALCIUM-ACTIVATED, PHOSPHOLIPID-DEPENDENT,
 CC SERINE- AND THREONINE-SPECIFIC ENZYME.
 CC -!- FUNCTION: PKC IS ACTIVATED BY DIACYLGLYCEROL WHICH IN TURN
 CC PHOSPHORYLATES A RANGE OF CELLULAR PROTEINS. PKC ALSO SERVES AS
 CC THE RECEPTOR FOR PHORBOL ESTERS, A CLASS OF TUMOR PROMOTERS.
 CC -!- SIMILARITY: CONTAINS 2 ZINC-DEPENDENT PHORBOL-ESTER AND DAG
 CC BINDING DOMAINS.
 CC -!- SIMILARITY: CONTAINS 1 C2 DOMAIN.
 CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 CC PKC SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X07287; CAA30267.1; -
 DR EMBL; M13707; AAA41874.1; -
 DR EMBL; M55417; AAA41873.1; -
 DR EMBL; X67129; CAA47608.1; -
 DR EMBL; L28035; AAA39939.1; -
 DR PIR; A05105; KIRPGC.
 DR PIR; S29620; S29620.
 DR PIR; JN0548; JN0548.
 DR PDB; ITBN; 29-APR-98.
 DR MGD; MGI:97597; Prkcc.
 DR InterPro; IPR000008; C2.
 DR InterPro; IPR002219; DAG_PE-bind.
 DR InterPro; IPR0000961; Euk_pkinase.
 DR InterPro; IPR002290; Ser_thr_pkinase.
 DR Pfam; PF00168; C2; 1.
 DR Pfam; PF00130; DAG_PE-bind; 2.
 DR Pfam; PF00069; pkinase; 1.
 DR Pfam; PF00433; pkinase; C; 1.
 DR PRINTS; PR00360; C2DOMAIN.
 DR PRINTS; PR00008; DAGPEDOMAIN.
 DR SMART; SM00109; C1; 2.
 DR SMART; SM00239; C2; 1.
 DR SMART; SM00133; S_TK_X; 1.
 DR SMART; SM00220; S_TK; 1.
 DR PROSITE; PS00499; C2DOMAIN_1; 1.
 DR PROSITE; PS50004; C2DOMAIN_2; 1.
 DR PROSITE; PS00479; DAG_PE_BIND_DOM_1; 2.
 DR PROSITE; PS50081; DAG_PE_BIND_DOM_2; 2.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE; PS50010; PROTEIN_KINASE_ST; 1.
 DR PROSITE; PS00108; Repeat; ATP-binding; transferase;
 KW Calcium-binding; Repeat; ATP-binding; transferase;
 KW Serine/threonine-protein kinase; Phorbol-ester binding; zinc;
 KW Phosphorylation; 3D-structure.

FT DOMAIN 36 85 PHORBOL-ESTER AND DAG BINDING 1.
 FT DOMAIN 101 150 PHORBOL-ESTER AND DAG BINDING 2.
 FT DOMAIN 170 260 C2 DOMAIN.
 FT DOMAIN 351 614 PROTEIN KINASE.
 FT NP_BIND 357 365 ATP (BY SIMILARITY).
 FT BINDING 380 380 ATP (BY SIMILARITY).
 FT ACT_SITE 480 480 BY SIMILARITY.
 FT MOD_RES 648 648 PHOSPHORYLATION (AUTO-) (POTENTIAL).
 FT MOD_RES 655 655 PHOSPHORYLATION (AUTO-) (POTENTIAL).
 SQ SEQUENCE 697 AA; 78357 MW; 56E2F7A3B93042FF CRC64;

 Query Match 0.43; Score 8; DB 1; Length 697;
 Best Local Similarity 100.0%; Pred. No. 15;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 Qy 227 GKSVDWWS 234
 Db 533 GKSVDWWS 540
 |||||
 |||||

 RESULT 18
 KPCG_RABBIT
 ID KPCG_RABBIT STANDARD; PRT; 697 AA.
 AC P10829;
 DT 01-JUL-1989 (Rel. 11, Created)
 DT 01-JUL-1989 (Rel. 11, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Protein kinase C, gamma type (EC 2.7.1.-) (PKC-gamma) (Delta).
 GN PKCG.
 OS Oryctolagus cuniculus (Rabbit).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
 OX NCBI_TaxID=9986;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE-88241036; PubMed-2837282;
 RM Ohno S., Kawasaki H., Konno Y., Inagaki M., Hidaka H., Suzuki K.;
 RT "A fourth type of rabbit protein kinase C.";
 RL Biochemistry 27:2083-2087(1988).
 CC -!- FUNCTION: THIS IS A CALCIUM-ACTIVATED, PHOSPHOLIPID-DEPENDENT,
 CC SERINE- AND THREONINE-SPECIFIC ENZYME.
 CC -!- FUNCTION: PKC IS ACTIVATED BY DIACYLGLYCEROL WHICH IN TURN
 CC PHOSPHORYLATES A RANGE OF CELLULAR PROTEINS. PKC ALSO SERVES AS
 CC THE RECEPTOR FOR PHORBOL ESTERS, A CLASS OF TUMOR PROMOTERS.
 CC -!- SIMILARITY: CONTAINS 2 ZINC-DEPENDENT PHORBOL-ESTER AND DAG
 CC BINDING DOMAINS.
 CC -!- SIMILARITY: CONTAINS 1 C2 DOMAIN.
 CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 CC PKC SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M13338; AAA31449.1; -
 DR PIR; A28708; KIRBGC.
 DR HSPSP; P05697; ITBN.
 DR InterPro; IPR000008; C2.
 DR InterPro; IPR002219; DAG_PE-bind.
 DR InterPro; IPR000719; Euk_pkinase.
 DR InterPro; IPR000961; Pkinase_C.
 DR InterPro; IPR002290; Ser_thr_pkinase.
 DR Pfam; PF00168; C2; 1.
 DR Pfam; PF00130; DAG_PE-bind; 2.
 DR Pfam; PF00069; pkinase; 1.
 DR Pfam; PF00433; pkinase; C; 1.
 DR PRINTS; PR00360; C2DOMAIN.
 DR PRINTS; PR00008; DAGPEDOMAIN.

DR SMART; SM00109; C1; 2.
 DR SMART; SM00239; C2; 1.
 DR SMART; SM00133; S-TK_X; 1.
 DR SMART; SM00220; S-TK; 1.
 DR PROSITE; PS00499; C2_DOMAIN_1; 1.
 DR PROSITE; PS00499; C2_DOMAIN_2; 1.
 DR PROSITE; PS00479; DAG_PE_BIND_DOM_1; 2.
 DR PROSITE; PS00479; DAG_PE_BIND_DOM_2; 2.
 DR PROSITE; PS00081; DAG_PE_BIND_DOM_1; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_DOM; 1.
 DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
 KW Calcium-binding; Repeat; ATP-binding; Transferase;
 KW Serine/threonine-protein kinase; Phorbol-ester binding; Zinc;
 KW Phosphorylation.
 FT DOMAIN 36 85 PHORBOL-ESTER AND DAG BINDING 1.
 FT DOMAIN 101 150 PHORBOL-ESTER AND DAG BINDING 2.
 FT DOMAIN 170 260 C2 DOMAIN.
 FT DOMAIN 351 614 PROTEIN KINASE.
 FT NP_BIND 357 365 ATP (BY SIMILARITY).
 FT BINDING 380 380 ATP (BY SIMILARITY).
 FT ACT_SITE 480 480 BY SIMILARITY.
 FT MOD_RES 648 648 PHOSPHORYLATION (AUTO-) (POTENTIAL).
 FT MOD_RES 655 655 PHOSPHORYLATION (AUTO-) (POTENTIAL).
 SQ SEQUENCE 697 AA; 78371 MW; 925D22221F78E5BF CRC64;

Query Match 0.4%; Score 8; DB 1; Length 697;
 Best Local Similarity 100.0%; Pred. No. 15;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 227 GKSVDWWS 234
 DB 533 GKSVDWWS 540
 |||||

RESULT 19
 ADDB_MOUSE STANDARD; PRT; 725 AA.
 AC Q9QYB8; Q9QYB7; Q9JLE5; Q9CXE3;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Beta adducin (Erythrocyte adducin beta subunit) (Add97).
 GN AD22.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OA NCBI_taxid=10090;
 RN [1]
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
 RX MEDLINE=200609066; PubMed=10602987;
 RA Suriyapperuma S.P., Lozovatsky L., Cicciotte S.L., Peters L.L.,
 RA Gilligan D.M.;
 FT "The mouse adducin gene family: alternative splicing and chromosomal
 FT localization.";
 RL Mamm. Genome 11:16-23(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J;
 RX MEDLINE=20304505; PubMed=10845937;
 RA Muro A.F., Marro M.L., Gajovic S., Porro F., Luzzatto L.,
 RA Baralle F.E.;
 RT "Mild spherocytic hereditary elliptocytosis and altered levels of
 RT alpha- and gamma-adducins in beta-adducin-deficient mice.";
 RL Blood 95:3978-3985(2000).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Embryonic liver;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishii K., Kiyosawa H., Kondo S., Yamakawa I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,

RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Harsh G.,
 RA Blake J., Boifelli D., Bojunga N., Carninci P., de Bonaudo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzaletti J., Nombarts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
 RA Yushakov-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,
 RA Hayashizaki Y.;
 Functional annotation of a full-length mouse cDNA collection.;
 RT Nature 409:685-690(2001).
 RL CCL
 CC FUNCTION: MEMBRANE-CYTOSKELETON-ASSOCIATED PROTEIN THAT PROMOTES
 CC THE ASSEMBLY OF THE SPECTRIN-ACTIN NETWORK. BINDS TO CALMODULIN.
 CC CALMODULIN BINDS PREFERENTIALLY TO THE BETA SUBUNIT (BY
 CC SIMILARITY).
 CC SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT.
 CC ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS; 1 (SHOWN HERE) AND 2;
 CC ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC DOMAIN: EACH SUBUNIT IS COMPRISED OF THREE REGIONS: A NH2-TERMINAL
 CC PROTEASE-RESISTANT GLOBULAR HEAD REGION, A SHORT CONNECTING
 CC SUBDOMAIN, AND A PROTEASE-SENSITIVE TAIL REGION.
 CC SIMILARITY: BELONGS TO THE ALDOLASE CLASS II FAMILY. ADDUCIN
 CC SUBFAMILY.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

EMBL; AF100422; AAF24972.1; -
 DR EMBL; AF100423; AAF24973.1; -
 DR EMBL; AF189769; AAF29502.1; -
 DR EMBL; AK014496; BAB29395.1; -
 MGD; MGI:87919; Add2.
 DR InterPro; IPR001303; Aldolase_II.
 DR Pfam; PF00596; Aldolase_II; 1.
 DR Cytoskeleton; Membrane; Calmodulin-binding; Phosphorylation;
 KW Alternative splicing.
 FT DOMAIN 425 444
 FT DOMAIN 703 720
 FT MOD_RES 55 55
 FT MOD_RES 702 702
 FT MOD_RES 712 712
 FT VARSPLIC 532 562
 FT VARSPLIC 563 725
 FT CONFLICT 469 469
 FT CONFLICT 634 634
 FT CONFLICT 666 666
 FT CONFLICT 675 675
 SQ SEQUENCE 725 AA; 80641 MW; 479153AB1BD6C0DA CRC64;

Query Match 0.4%; Score 8; DB 1; Length 725;
 Best Local Similarity 100.0%; Pred. No. 15;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1346 LVVNGKEE 1353
 DB 646 LVVNGKEE 653
 |||||

RESULT 20

CONT_CHICK
 ID CONT_CHICK STANDARD; PRT: 1010 AA.
 AC P14781; P10450;
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Contractin precursor (Neural cell recognition molecule F11).
 OS Gallus gallus (Chicken).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 CC Gallus.
 CC NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90180453; PubMed=2627374;
 RA Bruemendorf T., Wolff J.M., Rainer F., Rathjer F.G.;
 RT "Neural cell recognition molecule F11: homology with fibronectin type
 RT IRI and immunoglobulin type C domains.";
 RL Neuron 2:1351-1361(1989).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-WHITE LEGHORN;
 RX MEDLINE=89008597; PubMed=3049624;
 RA Ranscht B., Dours M.T.;
 RT "Sequence of contractin, a 130-kD glycoprotein concentrated in areas
 RT of interneuronal contact, defines a new member of the immunoglobulin
 RT supergene family in the nervous system.";
 RL J. Cell Biol. 107:1561-1573(1988).
 RN [3]
 RP GPI-ANCHOR.
 RX MEDLINE=89286606; PubMed=2735929;
 RA Wolff J.M., Bruemendorf T., Rathjen F.G.;
 RT "Neural cell recognition molecule F11: membrane interaction by
 RT covalently attached phosphatidylinositol.";
 RL Biochem. Biophys. Res. Commun. 161:931-938(1989).
 CC -!- FUNCTION: MEDIATES CELL SURFACE INTERACTIONS DURING NERVOUS
 CC SYSTEM DEVELOPMENT.
 CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
 CC -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
 CC -!- SIMILARITY: CONTAINS 6 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
 CC -!- SIMILARITY: CONTAINS 4 FIBRONECTIN TYPE IRI-LIKE DOMAINS.
 CC -!- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN IN THE C-TERMINUS
 CC AND IS LONGER DUE TO A FRAMESHIFT.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X14877; CAA33018.1; .
 CC EMBL; Y00813; CAA68753.1; ALT_FRAME.
 CC PIR; J00094; J00094.
 CC PIR; S01998; S01998.
 CC HSP; P40189; IBOU.
 CC InterPro; IPR003961; FN_IIL.
 CC InterPro; IPR003006; IG_MHC.
 CC InterPro; IPR003598; IG_C2.
 CC InterPro; IPR003600; IG_like.
 CC Pfam; PF00041; fn3; 4.
 CC Pfam; PF00047; ig; 6.
 CC SMART; SM00050; FN3; 4.
 CC SMART; SM00410; IG_like; 1.
 CC SMART; SM00408; IGc2; 5.
 CC Immunoglobulin domain; Glycoprotein; Signal; GPI-anchor;
 CC Cell adhesion; Repeat.
 FT SIGNAL 1 19
 FT CHAIN 20 2
 FT PROPEP ? 1010
 FT DOMAIN 50 113
 FT DOMAIN 143 210
 FT DOMAIN 193 210
 FT DOMAIN 210 210

FT DOMAIN 247 308
 FT DOMAIN 336 389
 FT DOMAIN 420 482
 FT DOMAIN 510 581
 FT DOMAIN 593 599
 FT DOMAIN 600 701
 FT DOMAIN 702 803
 FT DOMAIN 804 900
 FT DOMAIN 901 996
 FT CARBOHYD 200 200
 FT CARBOHYD 249 249
 FT CARBOHYD 329 329
 FT CARBOHYD 448 448
 FT CARBOHYD 464 464
 FT CARBOHYD 485 485
 FT CARBOHYD 512 512
 FT CARBOHYD 582 582
 FT CARBOHYD 924 924
 SQ SEQUENCE 1010 AA; 112507 MW; 2E38F071AF423AE1 CRC64;
 Query Match 0.4%; Score 8; DB 1; Length 1010;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1708 LKNGVSFR 1715
 DB 356 LKNGVSFR 363
 RESULT 21
 H1PR_MOUSE
 ID H1PR_MOUSE STANDARD; PRT: 1068 AA.
 AC Q9JKY5;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Huntingtin interacting protein 1 related (H1p1-related).
 GN H1P1.
 OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 CC NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=2008288; PubMed=10613908;
 RA Engqvist-Goldstein A.E.Y., Kessels M.M., Chopra V.S., Hayden M.R.,
 RA Drubin D.G.;
 RT "An actin-binding protein of the Sla2/Huntingtin interacting protein 1
 RT family is a novel component of clathrin-coated pits and vesicles.";
 RL J. Cell Biol. 147:1503-1518(1999).
 CC -!- FUNCTION: COMPONENT OF CLATHRIN-COATED PITS AND VESICLES, THAT MAY
 CC LINK THE ENDOCYTIC MACHINERY TO THE ACTIN CYTOSKELETON.
 CC -!- SUBUNIT: INTERACTS WITH ACTIN AND HUNTINGTIN INTERACTING PROTEIN 1
 CC (H1P1). DOES NOT INTERACT WITH HUNTINGTIN (BY SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC. MEMBRANE-ASSOCIATED PROTEIN,
 CC MAINLY LOCALIZED AT THE ENDOCYTIC COMPARTMENTS AND IN THE
 CC PERINUCLEAR REGION.
 CC -!- TISSUE SPECIFICITY: WIDELY EXPRESSED. EXPRESSED AT LOWER LEVELS IN
 CC SKELETAL MUSCLE AND HEART. THE LEVEL OF EXPRESSION DOES NOT CHANGE
 CC APPRECIABLY DURING DEVELOPMENT.
 CC -!- DOMAIN: THE TALIN-LIKE DOMAIN BINDS F-ACTIN.
 CC -!- SIMILARITY: BELONGS TO THE SLA2 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

```

DR EMBL; AF224713; AAF34662.1; -.
DR MGD; MGI:1352504; HipLr.
DR InterPro; IPR001026; ENTH.
DR InterPro; IPR002558; ILWEQ.
DR Pfam; PF01417; ENTH; 1.
DR Pfam; PF01608; ILWEQ; 1.
DR ProDom; PD011820; ILWEQ; 1.
DR SMART; SM00273; ENTH; 1.
DR SMART; SM00307; ILWEQ; 1.
KW Actin-binding; Coiled coil.
FT DOMAIN 77 80 POLY-SER.
FT DOMAIN 311 316 POLY-GLU.
FT DOMAIN 346 644 COILED COIL (POTENTIAL).
FT DOMAIN 800 1088 PALIN-LIKE.
SQ SEQUENCE 1068 AA; 119484 MW; 97CEE9D92CDF5DB1 CRC64;

Query Match 0.4%; Score 8; DB 1; Length 1068;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1293 DKVVLHMG 1300
DB 891 DKVVLHMG 898

RESULT 22
LHA2_RHOSH STANDARD; PRT; 54 AA.
AC P02946;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Light-harvesting protein B-800/850, alpha chain (LH-2) (Antenna
DE pigment protein, alpha chain).
GN PUCA.
OS Rhodobacter sphaeroides (Rhodospseudomonas sphaeroides).
OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
OC Rhodobacter.
OX NCBI_TaxId=1063;
RX MEDLINE-87250299; PubMed=3036782;
RA Kiley P.J., Kaplan S.;
RT "Cloning, DNA sequence, and expression of the Rhodobacter sphaeroides
RT light-harvesting B800-850-alpha and B800-850-beta genes.";
RL J. Bacteriol. 169:3268-3275(1987).
RN [2]
SEQUENCE FROM N.A.
RA Ashby M.K., Coomber S.A., Hunter C.N.;
RT "Cloning, nucleotide sequence and transfer of genes for the B800-850
RT light harvesting complex of Rhodobacter sphaeroides.";
RL FEBS Lett. 213:245-248(1987).
RN [3]
SEQUENCE FROM N.A.
RA MEDLINE-90132586; PubMed=2693605;
RA Burgess J.G., Ashby M.K., Hunter C.N.;
RT "Characterization and complementation of a mutant of Rhodobacter
RT sphaeroides with a chromosomal deletion in the light-harvesting (LH2)
RT genes.";
RL J. Gen. Microbiol. 135:1809-1816(1989).
RN [4]
SEQUENCE FROM N.A.
RC STRAIN=ATCC 17023 / 2.4.1 / NCIB 8253 / DSM 158;
RX MEDLINE-93086425; PubMed=1453956;
RA Gibson L.C., McGlynn P., Chaudhri M., Hunter C.N.;
RT "A putative anaerobic coproporphyrinogen III oxidase in Rhodobacter
RT sphaeroides. II. Analysis of a region of the genome encoding hemf and
RT the puc operon.";
RL Mol. Microbiol. 6:3171-3186(1992).
RN [5]
SEQUENCE FROM N.A.
RC STRAIN=ATCC 17023 / 2.4.1 / NCIB 8253 / DSM 158;

```

```

RX MEDLINE-20115911; PubMed=10648776;
RA Choudhary M., Kaplan S.;
RT "DNA sequence analysis of the photosynthesis region of Rhodobacter
RT sphaeroides 2.4.1.";
RL Nucleic Acids Res. 28:862-867(2000).
RN [6]
RP MUTANT R-26.1 SEQUENCE.
RX MEDLINE-85005243; PubMed=6384009;
RA Theiler R., Suter F., Wiemken V., Zuber H.;
RT "The light-harvesting polypeptides of Rhodospseudomonas sphaeroides R-
RT 26.1. I. Isolation, purification and sequence analyses.";
RL Hoppe-Sevler's Z. Physiol. Chem. 365:703-719(1984).
CC -1- FUNCTION: ANTENNA COMPLEXES ARE LIGHT-HARVESTING SYSTEMS, WHICH
CC TRANSFER THE EXCITATION ENERGY TO THE REACTION CENTERS.
CC -1- SUBUNIT: THE CORE COMPLEX IS FORMED BY DIFFERENT ALPHA AND BETA
CC CHAINS, BINDING BACTERIOCHLOROPHYLL MOLECULES, AND ARRANGED
CC MOST PROBABLY IN TETRAMERIC STRUCTURES DISPOSED AROUND THE
CC REACTION CENTER. THE NON-PIGMENTED GAMMA CHAINS MAY CONSTITUTE
CC ADDITIONAL COMPONENTS.
CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Inner membrane.
CC -1- MISCELLANEOUS: THIS POLYPEPTIDE (LH-2) AND LH-3B CONSTITUTE THE
CC B-800/850 COMPLEX OF R.SPHEROIDES 2.4.1 AND THE SPECTRALLY
CC ALTERED B-850 COMPLEX ISOLATED FROM THE BLUE-GREEN MUTANT R-26.1,
CC WHICH ABSORBS AT 860 NM.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M16777; AAA26160.1; -.
DR EMBL; X05200; CAA28833.1; -.
DR EMBL; M28360; AAA26131.1; -.
DR EMBL; X68796; CAA48700.1; -.
DR EMBL; AF195122; AAF24246.1; -.
DR PIR; A03446; LBRF2S.
DR PIR; A27087; A27087.
DR PIR; B26645; B26645.
DR PIR; B45795; B45795.
DR PIR; S28024; S28024.
DR HSP; P26789; LK2U.
DR InterPro; IPR002361; Antenna_comp_alpha.
DR InterPro; IPR000066; LHC.
DR Pfam; PF00556; LHC; 1.
DR PROSITE; PS00968; ANTENNA_COMP_ALPHA; 1.
KW Antenna complex; Light-harvesting polypeptide; Transmembrane;
KW Magnesium; Bacteriochlorophyll; Inner membrane.
FT DOMAIN 1 14 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 15 35 POTENTIAL.
FT DOMAIN 36 54 PERIPLASMIC (POTENTIAL).
FT METAL 31 31 AXIAL LIGAND TO THE BACTERIOCHLOROPHYLL
FT VARIANT 24 24 MAGNESIUM (POTENTIAL).
FT SEQUENCE 54 AA; 5599 MW; 50495C5332R23220 CRC64;
V -> F (IN MUTANT).

Query Match 0.4%; Score 7; DB 1; Length 54;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1310 AAVLTFT 1316
DB 32 AAVLTFT 38
|||||
|111111|

RESULT 23
RS19_BACSU STANDARD; PRT; 91 AA.
ID RS19_BACSU
AC P21476;
DT 01-MAY-1991 (Rel. 18, Created)

```

DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DE 16-OCT-2001 (Rel. 40, Last annotation update)
 DE 30S ribosomal protein S19 (BS19).
 GN RPSS.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RA Yoshikawa H., Yasumoto K., Takahashi H.;
 RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SG38;
 RA Li X., Lindahl L., Zengel J.M.;
 RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 1-35.
 RX MEDLINE=82219212; PubMed=6806564;
 RA Higo K.I., Otake E., Osawa S.;
 RT "Purification and characterization of 30S ribosomal proteins from
 RT Bacillus subtilis; correlation to Escherichia coli 30S proteins.";
 RL Mol. Gen. Genet. 185:239-244(1982).
 CC -!- FUNCTION: PROTEIN S19 FORMS A COMPLEX WITH S13 THAT BINDS STRONGLY
 CC TO THE 16S RIBOSOMAL RNA.
 CC -!- SIMILARITY: BELONGS TO THE S19P FAMILY OF RIBOSOMAL PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; D50302; BAA08835.1; -;
 CC EMBL; U43929; AAC45960.1; -;
 CC EMBL; Z99104; CAB11896.1; -;
 CC PIR; S11369; S11369.
 CC HSP; P80381; LQKF.
 CC Subtilist; BG19011; rpss.
 CC InterPro; IPR002222; Ribosomal_S19.
 CC Pfam; PF00203; Ribosomal_S19; 1.
 CC PRINTS; PR00975; RIBOSOMALS19.
 CC ProDom; PD001012; Ribosomal_S19; 1.
 CC PROSITE; PS00323; RIBOSOMAL_S19; 1.
 CC Ribosomal protein; rRNA-binding; Complete proteome.
 FT INIT_MET 0 0
 FT CONFLICT 11 11 D -> N (IN REF. 3).
 SQ SEQUENCE 91 AA; 10452 MW; 02BE15F85A/CEC89 CRC64;
 Query Match 0.4%; Score 7; DB 1; Length 91;
 Best Local Similarity 100.0%; Pred. No. 28;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 76 TDRKQV 82
 Db 24 TDRKQV 30
 RESULT 24
 IL8_PIG STANDARD; PRT; 103 AA.
 AC P26894; P22951;
 DT 01-AUG-1991 (Rel. 19, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Interleukin-8 precursor (IL-8) (Alveolar macrophage chemotactic factor
 DE I) (AMCF-I).
 GN *IL8.

OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 OX NCBI_TaxID=9823;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94103307; PubMed=8276881;
 RA Lin G., Pearson A.E., Scamurra R.W., Zhou Y., Baarsch M.J.,
 RA Weiss D.J., Murtaugh M.P.;
 RT "Regulation of interleukin-8 expression in porcine alveolar
 RT macrophages by bacterial lipopolysaccharide.";
 RL J. Biol. Chem. 269:77-85(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Sanjanwala M.;
 RL Submitted (JUL-1991) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 26-45.
 RC TISSUE=Lung;
 RX MEDLINE=93041741; PubMed=1420165;
 RA Goodman R.B., Foster D.C., Mathewes S.L., Osborn S.G., Kuijper J.L.,
 RA Forstrom J.W., Martin T.R.;
 RT "Molecular cloning of porcine alveolar macrophage-derived neutrophil
 RT chemotactic factors I and II; identification of porcine IL-8 and
 RT another intercrine-alpha protein.";
 RL Biochemistry 31:10483-10490(1992).
 RN [4]
 RP REVISION TO 23.
 RA Goodman R.B.;
 RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE OF 26-45.
 RC STRAIN=YORKSHIRE;
 RX MEDLINE=91217086; PubMed=1850745;
 RA Goodman R.B., Forstrom J.W., Osborn S.G., Chi E.Y., Martin T.R.;
 RT "Identification of two neutrophil chemotactic peptides produced by
 RT porcine alveolar macrophages.";
 RL J. Biol. Chem. 266:8455-8463(1991).
 CC -!- FUNCTION: IL-8 IS A CHEMOTACTIC FACTOR THAT ATTRACTS NEUTROPHILS,
 CC BASOPHILS, AND T-CELLS, BUT NOT MONOCYTES. IT IS ALSO INVOLVED IN
 CC NEUTROPHIL ACTIVATION. IT IS RELEASED FROM SEVERAL CELL TYPES IN
 CC RESPONSE TO AN INFLAMMATORY STIMULUS.
 CC -!- SUBUNIT: HOMODIMER.
 CC -!- TISSUE SPECIFICITY: ALVEOLAR MACROPHAGES.
 CC -!- INDUCTION: BY LIPOPOLYSACCHARIDE (LPS).
 CC -!- SIMILARITY: BELONGS TO THE INTERCRINE ALPHA FAMILY (SMALL CYTOKINE
 CC C-X-C) (CHEMOKINE CXC).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M86923; AAA16616.1; -;
 CC EMBL; X61151; CAA43461.1; -;
 CC EMBL; M99367; AAA92576.1; -;
 CC PIR; A44253; A44253.
 CC PIR; A39819; A39819.
 CC HSP; P10145; 1IKM.
 CC InterPro; IPR001811; Chemokine_IL8.
 CC InterPro; IPR001089; Small_cytokine_CXC.
 CC Pfam; PF00048; IL8; 1.
 CC PRINTS; PR00437; SMALLCYTCKXC.
 CC SMART; SM00199; SCY; 1.
 CC PROSITE; PS00471; SMALL_CYTOKINES_CXC; 1.
 KW Cytokine; Chemotaxis; Inflammatory response; Signal.
 FT SIGNAL 1 25
 FT CHAIN 26 103 INTERLEUKIN-8.
 FT DISULFID 34 61 BY SIMILARITY.
 FT DISULFID 36 77 BY SIMILARITY.


```

FT CONFLICT 33 34 RC -> CR (IN REF. 5).
SI CONFLICT 87 87 K -> KK (IN REF. 2).
SQ SEQUENCE 103 AA; 11633 MW; 9FE0E350E1928C64 CRC64;

Query Match
Best Local Similarity 0.4%; Score 7; DB 1; Length 103;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 639 LVNGREV 645
DB 70 LVNGREV 76

RESULT 25
YB15_HALN1
ID YB15_HALN1 STANDARD; PRT: 124 AA.
AC Q9HOK7;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein Vng1115H.
GN Vng1115H.
OS Halobacterium sp. (strain NRC-1).
OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;
OC Halobacterium.
OX NCBI_TaxID=64091;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20504483; PubMed=11016950;
RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,
RA Swartzell S., Weir D., Hall J., Dahl T.A., Weir R., Goo Y.A.,
RA Leithausen B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA Ebbardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.;
RA "Genome sequence of Halobacterium species NRC-1";
RA Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
CC -!- SIMILARITY: BELONGS TO THE UPF0066 (VIRR) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AJ245645; CAB53114.1; -
CC InterPro; IPR001857; Ribosomal_L19.
CC Pfam; PF01245; Ribosomal_L19; 1.
CC PRINTS; PR00061; RIBOSOMAL_L19.
CC PRODOM; PD002979; Ribosomal_L19; 1.
CC PROSITE; PS01015; RIBOSOMAL_L19; 1.
CC Ribosomal protein; Chloroplast.
CC SEQUENCE 129 AA; 14793 MW; 444FAD8CB1DDA1D CRC64;

Query Match
Best Local Similarity 0.4%; Score 7; DB 1; Length 129;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1498 ALHKAGL 1504
DB 52 ALHKAGL 58

RESULT 27
YQ09_HALN1
ID YQ09_HALN1 STANDARD; PRT: 129 AA.
AC P58015;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein Vng2609C.
GN Vng2609C.
OS Halobacterium sp. (strain NRC-1).
OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;
OC Halobacterium.
OX NCBI_TaxID=64091;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20504483; PubMed=11016950;
RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,
RA Swartzell S., Weir D., Hall J., Dahl T.A., Weir R., Goo Y.A.,
RA Leithausen B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA Ebbardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.;
RA "Genome sequence of Halobacterium species NRC-1";
RA Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
CC -!- SIMILARITY: BELONGS TO THE UPF0066 (VIRR) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE005042; AAG19506.1; -
CC InterPro; IPR001378; UPF0066.
CC Pfam; PF01980; UPF0066; 1.
CC PROSITE; PS01318; UPF0066; 1.
CC Hypothetical protein; Complete proteome.
CC SEQUENCE 124 AA; 13302 MW; 504BEFF7634CALC6 CRC64;

Query Match
Best Local Similarity 0.4%; Score 7; DB 1; Length 124;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1713 SPRGGLV 1719
DB 37 SPRGGLV 43

RESULT 26
RK19_PROWI
ID RK19_PROWI STANDARD; PRT: 129 AA.
AC Q9TUG7;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)

```

```

DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Chloroplast 50S ribosomal protein L19.
GN RPL19.
OS Chlorotheca wickerhamii.
OC Chloroplast.
OC Eukaryota; Viridiplantae; Chlorophyta; Trebouxiophyceae; Chlorellales;
OC Chlorallaceae; Chlorotheca.
OX NCBI_TaxID=3111;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=263-11;
RA Knauf U., Hachtel W.;
RA "A 22 kb fragment of the 53 kb plastid genome of the colourless alga
RA Chlorotheca wickerhamii containing atp-, rpl-, rps-, rrm-, and trn-
RA genes.";
RA Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE L19P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AJ245645; CAB53114.1; -
CC InterPro; IPR001857; Ribosomal_L19.
CC Pfam; PF01245; Ribosomal_L19; 1.
CC PRINTS; PR00061; RIBOSOMAL_L19.
CC PRODOM; PD002979; Ribosomal_L19; 1.
CC PROSITE; PS01015; RIBOSOMAL_L19; 1.
CC Ribosomal protein; Chloroplast.
CC SEQUENCE 129 AA; 14793 MW; 444FAD8CB1DDA1D CRC64;

Query Match
Best Local Similarity 0.4%; Score 7; DB 1; Length 129;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1498 ALHKAGL 1504
DB 52 ALHKAGL 58

RESULT 27
YQ09_HALN1
ID YQ09_HALN1 STANDARD; PRT: 129 AA.
AC P58015;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein Vng2609C.
GN Vng2609C.
OS Halobacterium sp. (strain NRC-1).
OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;
OC Halobacterium.
OX NCBI_TaxID=64091;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20504483; PubMed=11016950;
RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,
RA Swartzell S., Weir D., Hall J., Dahl T.A., Weir R., Goo Y.A.,
RA Leithausen B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA Ebbardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.;
RA "Genome sequence of Halobacterium species NRC-1";
RA Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
CC -!- SIMILARITY: BELONGS TO THE UPF0066 (VIRR) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE005042; AAG19506.1; -
CC InterPro; IPR001378; UPF0066.
CC Pfam; PF01980; UPF0066; 1.
CC PROSITE; PS01318; UPF0066; 1.
CC Hypothetical protein; Complete proteome.
CC SEQUENCE 124 AA; 13302 MW; 504BEFF7634CALC6 CRC64;

Query Match
Best Local Similarity 0.4%; Score 7; DB 1; Length 124;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1713 SPRGGLV 1719
DB 37 SPRGGLV 43

RESULT 26
RK19_PROWI
ID RK19_PROWI STANDARD; PRT: 129 AA.
AC Q9TUG7;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)

```

RA Bairoch A.;
 CC Unpublished observations (APR-2001).
 CC -|- SIMILARITY: BELONGS TO THE UPF0146 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AE005134; -; NOT_ANNOTATED_CDS.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 129 AA; 13289 MW; 0C68CFD215373530 CRC64;

 Query Match 0.4%; Score 7; DB 1; Length 129;
 Best Local Similarity 100.0%; Pred. No. 37;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 535 ARPDVAA 541
 Db 24 ARPDVAA 30
 |||||

 RESULT 28
 ID Y996_ARCFU STANDARD; PRT; 141 AA.
 AC Q29169;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein Af1096.
 GN AF1096.
 OS Archaeoglobus fulgidus.
 CC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
 CC Archaeoglobus.
 CC NCBI_TaxID=2234;
 [1]
 SEQUENCE FROM N.A.
 RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
 RX MEDLINE=98049343; PubMed=9389475;
 RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
 RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
 RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,
 RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
 RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
 RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
 RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Uitterback T.,
 RA Cotton M.D., Spriggs T., Artlich P., Kaine B.P., Sykes S.M.,
 RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
 RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
 RA Venter J.C.;
 RT "The complete genome sequence of the hyperthermophilic, sulphate-
 RT reducing archaeon Archaeoglobus fulgidus."
 RL Nature 390:364-370(1997).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AE001028; AAB90161.1; -;
 DR TIGR; AF1096; -;
 KW Hypothetical protein; Transmembrane; Complete proteome.
 FT TRANSMEM 13 35 POTENTIAL.
 SQ SEQUENCE 141 AA; 14586 MW; 8809C8ACB0F8C16E CRC64;

Query Match 0.4%; Score 7; DB 1; Length 141;
 Best Local Similarity 100.0%; Pred. No. 40;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 638 VLVNGKE 644
 Db 79 VLVNGKE 85
 |||||

 RESULT 29
 ID RP88_YEAST STANDARD; PRT; 146 AA.
 AC P20436;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE DNA-directed RNA polymerases I, II, and III 14.5 kDa polypeptide
 DE (EC 2.7.7.6) (AEC14.4).
 GN RP88 OR YOR224C OR YOR50-14.
 OS Saccharomyces cerevisiae (Baker's yeast).
 CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 CC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 CC NCBI_TaxID=4932;
 [1]
 SEQUENCE FROM N.A.
 RX MEDLINE=90249736; PubMed=2186966;
 RA Woychik N.A., Liao S.-M., Kolodziej P.A., Young R.A.;
 RT "Subunits shared by eukaryotic nuclear RNA polymerases.";
 RL Genes Dev. 4:313-323(1990).
 [2]
 SEQUENCE FROM N.A.
 RC STRAIN=S288C / FY1679;
 RX MEDLINE=96437977; PubMed=8840505;
 RA Gallison F., Dujon B.;
 RT "Sequence and analysis of a 33 kb fragment from the right arm of
 RT chromosome XV of the yeast Saccharomyces cerevisiae.";
 RL Yeast 12:877-885(1996).
 [3]
 RP STRUCTURE BY NMR.
 RX MEDLINE=98120973; PubMed=9461075;
 RA Krapp S., Kelly G., Reischl J., Weinzierl R.O.J., Matthews S.;
 RT "Eukaryotic RNA polymerase subunit RP88 is a new relative of the OB
 RT family.";
 RL Nat. Struct. Biol. 5:110-114(1998).
 CC -|- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
 CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
 CC SUBSTRATES.
 CC -|- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +
 CC (RNA)(N).
 CC -|- SUBUNIT: EACH CLASS OF RNA POLYMERASE IS ASSEMBLED FROM 9 TO 15
 CC DIFFERENT POLYPEPTIDES. THIS SUBUNIT IS SHARED BY ALL 3 YEAST RNA
 CC POLYMERASES.
 CC -|- SUBCELLULAR LOCATION: Nuclear.
 CC -|- MISCELLANEOUS: THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES ARE
 CC FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA
 CC PRECURSOR, POLYMERASE II FOR THE MRNA PRECURSOR, AND POLYMERASE
 CC III FOR 5S AND TRNA GENES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X53289; CAA37383.1; -;
 DR EMBL; X92441; CAA63187.1; -;
 DR EMBL; Z75132; CAA99443.1; -;
 DR PIR; C34588; C34588.
 DR PIR; S31227; S31227.
 DR PDB; 1AID; 02-MAR-99.
 DR SGD; S0005750; RP88.

KW Transferase; DNA-directed RNA polymerase; Transcription;
KW Nuclear protein; 3D-structure.
SQ SEQUENCE 146 AA; 16511 MW; 5B71693F86FD863 CRC64;

Query Match 0.48; Score 7; DB 1; Length 146;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 45 ASTTQDQ 51
|||||
Db 29 ASTTQDQ 35

RESULT 30
RISB_STRCO STANDARD; PRT: 161 AA.
AC Q9EWJ9;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE 6,7-dimethyl-8-ribityllumazine synthase (EC 2.5.1.9) (DMRL synthase)
DE (Lumazine synthase) (Riboflavin synthase beta chain).
GN RIBH OR SC6D7A.03C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Harris D., Cerdano A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (NOV-2000) to the EMBL/Genbank/DBJ databases.
CC -!- FUNCTION: Riboflavin synthase is a bifunctional enzyme complex
CC catalyzing the formation of riboflavin from 5-amino-6-(1'-D)-
CC ribityl-amino-2,4(1H,3H)-pyrimidinedione and L-3,4-dihydroxy-2-
CC butanone-4-phosphate via 6,7-dimethyl-8-lumazine. The beta subunit
CC catalyzes the condensation of 5-amino-6-(1'-D)-ribityl-amino-
CC 2,4(1H,3H)-pyrimidinedione with L-3,4-dihydroxy-2-butanone-4-
CC phosphate yielding 6,7-dimethyl-8-lumazine (by similarity).
CC -!- CATALYTIC ACTIVITY: 2 6,7-dimethyl-8-(1-D-ribityl)lumazine =
CC riboflavin + 4-(1-D-ribitylamino)-5-amino-2,6-dihydroxypyrimidine.
CC -!- PATHWAY: Riboflavin synthesis; last step.
CC -!- SIMILARITY: BELONGS TO THE DMRL SYNTHASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AL450443; CAC17558.1; -
CC InterPro: IPR002180; DMRL_synthase.
CC Pfam: PF00885; DMRL_synthase; 1.
CC Riboflavin biosynthesis; Transference.
SQ SEQUENCE 161 AA; 16785 MW; 75662F87ABFB4CE CRC64;

Query Match 0.48; Score 7; DB 1; Length 161;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 17 VTAAYAT 23
|||||
Db 143 VTAAYAT 149

Search completed: August 12, 2002, 08:15:03
Job time: 309 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 12, 2002, 08:10:19 ; Search time 60.41 Seconds
(without alignments)
5100.215 Million cell updates/sec

Title: US-09-995-749a-2

Perfect score: 1781

Sequence: 1 MEIKKFKLYKSGKQWTA.....SDAEYPTSDVGKMLDQNK 1781

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 562222 seqs, 172994929 residues

Word size : 0

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database :

SPTREMBL19:*

1: sp.archaea:*

2: sp.bacteria:*

3: sp.fungi:*

4: sp.human:*

5: sp.invertebrate:*

6: sp.mammal:*

7: sp.mhc:*

8: sp.organelle:*

9: sp.phage:*

10: sp.plant:*

11: sp.protist:*

12: sp.virus:*

13: sp.vertibrate:*

14: sp.unclassified:*

15: sp.rviro:*

16: sp.bacteriap:*

17: sp.archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	17	1.0	1599	2 Q00599	Q00599 streptococc
2	16	0.9	1575	2 Q9LCH3	Q9LCH3 streptococc
3	16	0.9	1577	2 Q34178	Q34178 streptococc
4	16	0.9	1577	2 Q5265	Q5265 streptococc
5	15	0.8	1016	2 Q9LCH7	Q9LCH7 leuconostoc
6	15	0.8	1290	2 Q48756	Q48756 leuconostoc
7	15	0.8	1390	2 Q69385	Q69385 streptococc
8	15	0.8	1455	2 Q69382	Q69382 streptococc
9	15	0.8	1455	2 Q69388	Q69388 streptococc
10	15	0.8	1455	2 Q69391	Q69391 streptococc
11	15	0.8	1455	2 Q69397	Q69397 streptococc
12	15	0.8	1477	2 Q9L466	Q9L466 leuconostoc
13	15	0.8	1508	2 Q9E2H5	Q9E2H5 leuconostoc
14	15	0.8	1590	2 Q99983	Q99983 streptococc
15	15	0.8	1590	2 Q5263	Q5263 streptococc
16	13	0.7	1449	2 Q68542	Q68542 streptococc

17	13	0.7	1449	2 Q55264	Q55264 streptococc
18	13	0.7	1508	2 Q52224	Q52224 leuconostoc
19	13	0.7	1518	2 Q00600	Q00600 streptococc
20	13	0.7	1527	2 Q9ZAR4	Q9ZAR4 leuconostoc
21	13	0.7	2057	2 Q9RE05	Q9RE05 leuconostoc
22	11	0.6	1338	2 Q9WXJ4	Q9WXJ4 streptococc
23	10	0.6	1512	2 Q9WXJ5	Q9WXJ5 streptococc
24	9	0.5	20	2 Q9W4L7	Q9W4L7 leuconostoc
25	8	0.4	100	6 Q18799	Q18799 gorilla gor
26	8	0.4	123	6 Q18797	Q18797 gorilla gor
27	8	0.4	196	16 Q9CNV9	Q9CNV9 pasteurella
28	8	0.4	213	2 Q9KYA1	Q9KYA1 streptomyce
29	8	0.4	228	16 Q53780	Q53780 mycobacteri
30	8	0.4	292	10 Q65770	Q65770 craterostig
31	8	0.4	297	16 Q928V2	Q928V2 chlamydia p
32	8	0.4	297	16 Q9JS15	Q9JS15 chlamydia p
33	8	0.4	304	16 Q9AB26	Q9AB26 caulobacter
34	8	0.4	312	2 Q9L103	Q9L103 streptomyce
35	8	0.4	330	2 Q55228	Q55228 streptococc
36	8	0.4	334	3 Q9P8F7	Q9P8F7 yarrowia l1
37	8	0.4	338	5 Q9VW96	Q9VW96 drosophila
38	8	0.4	355	16 Q92B72	Q92B72 listeria in
39	8	0.4	383	10 Q94OR0	Q94OR0 arabidopsis
40	8	0.4	399	17 Q9HJR3	Q9HJR3 thermoplasm
41	8	0.4	442	10 Q9FED7	Q9FED7 oryza sativ
42	8	0.4	458	16 P71004	P71004 bacillus su
43	8	0.4	477	11 Q9JLE4	Q9JLE4 mus musculu
44	8	0.4	489	3 Q74505	Q74505 schizosacch
45	8	0.4	508	16 Q9M666	Q9M666 rhizobium l
46	8	0.4	561	10 Q9C7N2	Q9C7N2 arabidopsis
47	8	0.4	573	12 Q9EN10	Q9EN10 amsacta moo
48	8	0.4	626	4 Q9H967	Q9H967 homo sapien
49	8	0.4	629	5 Q97065	Q97065 drosophila
50	8	0.4	644	17 Q974G9	Q974G9 sulfolobus
51	8	0.4	777	16 Q97S78	Q97S78 streptococc
52	8	0.4	830	2 Q9AGP8	Q9AGP8 arthrobacte
53	8	0.4	1068	11 Q9D689	Q9D689 mus musculu
54	8	0.4	1079	11 Q9DPW9	Q9DPW9 rattus norv
55	8	0.4	1136	11 Q04692	Q04692 mus musculu
56	8	0.4	1331	10 Q49549	Q49549 arabidopsis
57	8	0.4	1495	10 Q9SUA7	Q9SUA7 arabidopsis
58	8	0.4	1591	10 Q9LDW9	Q9LDW9 oryza sativ
59	8	0.4	2265	5 Q9XV64	Q9XV64 caenorhabdi
60	8	0.4	3263	16 Q9KL97	Q9KL97 vibrio chol
61	8	0.4	3591	2 Q9K1E1	Q9K1E1 streptomyce
62	7	0.4	42	12 Q09525	Q09525 classical s
63	7	0.4	50	5 Q26209	Q26209 plasmodium
64	7	0.4	50	5 Q26214	Q26214 plasmodium
65	7	0.4	50	5 Q26215	Q26215 plasmodium
66	7	0.4	73	11 Q88742	Q88742 mus musculu
67	7	0.4	85	5 Q9VAZ9	Q9VAZ9 drosophila
68	7	0.4	101	6 Q9XSX5	Q9XSX5 felis silve
69	7	0.4	101	16 Q92G82	Q92G82 rickettsia
70	7	0.4	105	5 Q20197	Q20197 caenorhabdi
71	7	0.4	108	2 Q52420	Q52420 pseudomonas
72	7	0.4	108	3 Q93917	Q93917 pisolithus
73	7	0.4	110	12 Q9G6X6	Q9G6X6 heliocoverp
74	7	0.4	112	10 Q9LR88	Q9LR88 arabidopsis
75	7	0.4	115	10 Q9SNR6	Q9SNR6 oryza sativ
76	7	0.4	115	10 Q9SNR1	Q9SNR1 oryza sativ
77	7	0.4	119	16 Q97G55	Q97G55 ciostridium
78	7	0.4	126	10 Q93WH0	Q93WH0 arabidopsis
79	7	0.4	131	6 Q18847	Q18847 pongo pygma
80	7	0.4	132	6 Q18848	Q18848 hylolobus l
81	7	0.4	133	4 Q96AR0	Q96AR0 homo sapien
82	7	0.4	134	2 Q49963	Q49963 mycobacteri
83	7	0.4	134	10 Q9SUF2	Q9SUF2 arabidopsis
84	7	0.4	140	3 Q12160	Q12160 saccharomyc
85	7	0.4	144	16 Q9X0W8	Q9X0W8 thermotoga
86	7	0.4	144	16 Q97JW9	Q97JW9 ciostridium
87	7	0.4	149	2 Q9RJZ2	Q9RJZ2 streptomyce
88	7	0.4	150	12 Q11389	Q11389 rhesus papi
89	7	0.4	150	16 Q97HB2	Q97HB2 ciostridium

90 7 0.4 152 16 p73118 p73118 synchocyst
 91 7 0.4 156 2 Q9KY19 Q9KY19 streptomyce
 92 7 0.4 157 16 Q98110 Q98110 rhizobium 1
 93 7 0.4 158 5 Q9BJ46 Q9BJ46 leishmania
 94 7 0.4 161 2 Q93TSS Q93TSS uncultured
 95 7 0.4 161 17 Q9HJ41 Q9HJ41 thermoplasm
 96 7 0.4 162 16 Q9PG28 Q9PG28 xylella fas
 97 7 0.4 163 16 Q9K886 Q9K886 bacillus ha
 98 7 0.4 164 17 Q52033 Q52033 halobacteri
 99 7 0.4 166 5 Q9N7N4 Q9N7N4 leishmania
 100 7 0.4 166 17 Q9YCD6 Q9YCD6 aeropyrum p

ALIGNMENTS

RESULT 1
 ID Q00599 PRELIMINARY; PRT; 1599 AA.
 AC Q00599;
 DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
 DE GLUCOSYLTRANSFERASE S PRECURSOR (EC 2.4.1.5) (GTF) (DEXTRANSUCRASE)
 DE (SUCROSE 6-GLUCOSYLTRANSFERASE).
 GN GTFK.
 OS Streptococcus salivarius.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1304;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 25975;
 RX MEDLINE=93381463; PubMed=1838391;
 RA Giffard P.M., Simpson C.L., Milward C.P., Jacques N.A.;
 RT "Molecular characterization of a cluster of at least two
 glucosyltransferase genes in Streptococcus salivarius ATCC 25975.";
 RJ J. Gen. Microbiol. 137:2577-2593(1991).
 CC -!- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT TO
 PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE OF
 THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIANE THE
 AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.
 CC -!- CATALYTIC ACTIVITY: SUCROSE + ((1,6)-ALPHA-D-GLUCOSYL)(N) - D-
 FRUCTOSE + ((1,6)-ALPHA-D-GLUCOSYL)(N+1).
 CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
 CC -!- DISEASE: DENTAL CARIES.
 CC -!- SIMILARITY: TO REGIONS OF BAXLEY AND BACILLUS AMYLOLIQUEFACIENS
 ALPHA AMYLASES AND RABBIT GLYCOGEN PHOSPHORYLASE.
 DR EMBL; 211872; CAA77898.1; .
 DR EMBL; 211873; CAA77901.1; .
 DR EMBL; M64111; AAA26897.1; .
 DR InterPro; IPR002479; CW_binding.
 DR InterPro; IPR003318; Glyco_hydro_70.
 DR Pfam; PF01473; CW_binding_1; 14.
 DR Pfam; PF02324; Glyco_hydro_70; 1.
 KW Transferase; Glycosyltransferase; Signal; Repeat; Dental caries.
 FT SIGNAL 1 42 POTENTIAL
 FT CHAIN 43 1599 GLUCOSYLTRANSFERASE S.
 FT SIMILAR 867 967 TO ALPHA-AMYLASES.
 FT SIMILAR 1132 1219 TO GLYCOGEN PHOSPHORYLASE.
 SQ SEQUENCE 1599 AA; 176480 NW; 24577869E152B707 CRC64;

Query Match 1.08; Score 17; DB 2; Length 1599;
 Best Local Similarity 100.0%; Pred. No. 8.8e-08;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1560 YGGAFLDKLQKLYPEIF 1576
 Db 1011 YGGAFLDKLQKLYPEIF 1027
 RESULT 2
 ID Q9LCH3 PRELIMINARY; PRT; 1575 AA.
 AC Q9LCH3;
 DT 01-OCT-2000 (TREMblrel. 15, Created)
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
 DE GLUCOSYLTRANSFERASE.
 GN GTFK.
 OS Streptococcus oralis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1303;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC10557;
 RX MEDLINE=20231779; PubMed=10768934;
 RA Fujiwara T., Hoshino T., Ooshima T., Sobue S., Hamada S.;
 RT "Purification, characterization, and molecular analysis of the gene
 encoding glucosyltransferase from Streptococcus oralis.";
 RJ Infect. Immun. 68:2475-2483(2000).
 DR EMBL; AB025228; BAA95201.1; .
 DR InterPro; IPR002479; CW_binding.
 DR InterPro; IPR003318; Glyco_hydro_70.
 DR Pfam; PF01473; CW_binding_1; 17.
 DR Pfam; PF02324; Glyco_hydro_70; 1.
 KW Transferase.
 SQ SEQUENCE 1575 AA; 176792 NW; 772A26E4D7C2E543 CRC64;

Query Match 0.98; Score 16; DB 2; Length 1575;
 Best Local Similarity 100.0%; Pred. No. 9.4e-07;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1508 ADWVPDQIYNLPKREV 1523
 Db 1003 ADWVPDQIYNLPKREV 1018
 RESULT 3
 ID Q54178 PRELIMINARY; PRT; 1577 AA.
 AC Q54178; Q54247;
 DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
 DE GLUCOSYLTRANSFERASE.
 GN GTFK.
 OS Streptococcus gordonii challis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=29390;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CHALLIS;
 RX MEDLINE=96157084; PubMed=8586195;
 RA Vickerman M.M., Sulavik M.C., Clewell D.B.;
 RT "Molecular analysis of Streptococcus gordonii glucosyltransferase
 phase variants.";
 RJ Dev. Biol. Stand. 85:309-314(1995).
 RN [2]
 RP SEQUENCE OF 1-96 FROM N.A.
 RC STRAIN=CHALLIS;
 RX MEDLINE=92276337; PubMed=1534326;
 RA Sulavik M.C., Tardif G., Clewell D.B.;
 RT "Identification of a gene, rgg, which regulates expression of

```
RT glucosyltransferase and influences the Spp phenotype of Streptococcus
RT gordonii Challis.*;
RL J. Bacteriol. 174:3577-3586(1992).
DR EMBL; U12643; AAC43483.1; -.
DR EMBL; M89776; AAA26969.1; -.
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 18.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Transferase.
SQ SEQUENCE 1577 AA; 177805 MW; 5AE0328DC5E08D18 CRC64;

Query Match          0.9%; Score 16; DB 2; Length 1577;
Best Local Similarity 100.0%; Pred. No. 9.4e-07;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1508 ADWVPDQIYNLPKEV 1523
DB 1005 ADWVPDQIYNLPKEV 1020
|||||
RESULT 4
Q5265 PRELIMINARY; PRT; 1577 AA.
AC Q5265;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE GLUCOSYLTRANSFERASE PRECURSOR.
GN GTFM.
OS Streptococcus salivarius.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1304;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-95122197; PubMed-7822030;
RA Simpson C.L., Giffard P.M., Jacques N.A.;
RT "Streptococcus salivarius ATCC 25975 possesses at least two genes
RT coding for primer-independent glucosyltransferases.*";
RL Infect. Immun. 63:609-621(1995).
DR EMBL; I35928; AAC41413.1; -.
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 11.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Signal; transferase.
FT SIGNAL 1 38 POTENTIAL.
FT CHAIN 39 1577 GLUCOSYLTRANSFERASE.
SQ SEQUENCE 1577 AA; 175290 MW; 3EFB898A7D3A7BF3 CRC64;

Query Match          0.9%; Score 16; DB 2; Length 1577;
Best Local Similarity 100.0%; Pred. No. 9.4e-07;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1508 ADWVPDQIYNLPKEV 1523
DB 1035 ADWVPDQIYNLPKEV 1050
|||||
RESULT 5
Q9LCJ7 PRELIMINARY; PRT; 1016 AA.
AC Q9LCJ7;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE DEXTRANSUCRASE.
GN DSRT.
OS Leuconostoc mesenteroides.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
```

```
OC Leuconostoc.
OX NCBI_TaxID=1245;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-NRRL B-512F;
RX MEDLINE-20169623; PubMed-10705445;
RA Funane K., Mizuno K., Takahara H., Kobayashi M.;
RT *Gene encoding a dextranucrase-like protein in Leuconostoc
RT mesenteroides NRRL B-512F.*;
RL Biosci. Biotechnol. Biochem. 64:29-38(2000).
DR EMBL; AB020020; BAA90527.1; -.
DR HSP; P06278; 1VJS.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF02324; Glyco_hydro_70; 1.
SQ SEQUENCE 1016 AA; 110344 MW; 88968FDE13CCCB47 CRC64;

Query Match          0.8%; Score 15; DB 2; Length 1016;
Best Local Similarity 100.0%; Pred. No. 7e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1366 VSGYLAVVPVGASD 1380
DB 868 VSGYLAVVPVGASD 882
|||||
RESULT 6
Q48756 PRELIMINARY; PRT; 1290 AA.
AC Q48756;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE DEXTRANSUCRASE.
OS Leuconostoc mesenteroides.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
OX Leuconostoc.
OX NCBI_TaxID=1245;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-NRRL B1299;
RX MEDLINE-97136686; PubMed-8982063;
RA Monchois V., Willemot R.M., Renaud-Simeon M., Croux C., Monsan P.;
RT *Cloning and sequencing of a gene coding for a novel dextranucrase
RT from Leuconostoc mesenteroides NRRL B-1299 synthesizing only alpha (1-
RT 6) and alpha (1-3) linkages.*;
RL Gene 182:23-32(1996).
DR EMBL; U38181; AAB40875.1; -.
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 11.
DR Pfam; PF02324; Glyco_hydro_70; 1.
SQ SEQUENCE 1290 AA; 145590 MW; 3555C2E96B749PAA CRC64;

Query Match          0.8%; Score 15; DB 2; Length 1290;
Best Local Similarity 100.0%; Pred. No. 8.6e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 986 DNSNPVQAEQLNWL 1000
DB 248 DNSNPVQAEQLNWL 262
|||||
RESULT 7
O69385 PRELIMINARY; PRT; 1390 AA.
AC O69385;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE GLUCOSYLTRANSFERASE-SI.
GN GTFC.
```

OS Streptococcus mutans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
ON Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-WT4245;
RX MEDLINE-98231643; PubMed-9570124;
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,
Kimura S., Hamada S.;
RT "Molecular analyses of glucosyltransferase genes among strains of
Streptococcus mutans";
RL FEMS Microbiol. Lett. 161:331-336(1998).
DR EMBL; D88655; BAA26106.1; -;
DR InterPro; IPR002479; CW_binding.
DR Pfam; PF01473; Glyco_hydro_70.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Transference.
SQ SEQUENCE 1390 AA; 155375 MW; 8847E4956EF05E9F CRC64;

Query Match 0.8%; Score 15; DB 2; Length 1390;
Best Local Similarity 100.0%; Pred. No. 9.1e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 986 DNSNPVQAEQLNWL 1000
Db 439 DNSNPVQAEQLNWL 453

RESULT 8
O69382 PRELIMINARY; PRT; 1455 AA.
AC O69382;
DT 01-AUG-1998 (TRENBLrel. 07, Created)
DT 01-AUG-1998 (TRENBLrel. 07, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE GLUCOSYLTRANSFERASE-SI.
GN GTFC.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
ON Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-WT8148;
RX MEDLINE-98231643; PubMed-9570124;
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,
Kimura S., Hamada S.;
RT "Molecular analyses of glucosyltransferase genes among strains of
Streptococcus mutans";
RL FEMS Microbiol. Lett. 161:331-336(1998).
DR EMBL; D88652; BAA26102.1; -;
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 9.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Transference.
SQ SEQUENCE 1455 AA; 162970 MW; 27D4D3A1EBCA2039 CRC64;

Query Match 0.8%; Score 15; DB 2; Length 1455;
Best Local Similarity 100.0%; Pred. No. 9.5e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 986 DNSNPVQAEQLNWL 1000
Db 439 DNSNPVQAEQLNWL 453

RESULT 9
O69388 PRELIMINARY; PRT; 1455 AA.
AC O69388;
DT 01-AUG-1998 (TRENBLrel. 07, Created)
DT 01-AUG-1998 (TRENBLrel. 07, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE GLUCOSYLTRANSFERASE-SI.
GN GTFC.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
ON Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-WT4239;
RX MEDLINE-98231643; PubMed-9570124;
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,
Kimura S., Hamada S.;
RT "Molecular analyses of glucosyltransferase genes among strains of
Streptococcus mutans";
RL FEMS Microbiol. Lett. 161:331-336(1998).
DR EMBL; D88661; BAA26114.1; -;
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 9.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Transference.
SQ SEQUENCE 1455 AA; 162804 MW; 683A359D873E9E1A CRC64;

Query Match 0.8%; Score 15; DB 2; Length 1455;
Best Local Similarity 100.0%; Pred. No. 9.5e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

ID O69388 PRELIMINARY; PRT; 1455 AA.
AC O69388;
DT 01-AUG-1998 (TRENBLrel. 07, Created)
DT 01-AUG-1998 (TRENBLrel. 07, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE GLUCOSYLTRANSFERASE-SI.
GN GTFC.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
ON Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-WT4239;
RX MEDLINE-98231643; PubMed-9570124;
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,
Kimura S., Hamada S.;
RT "Molecular analyses of glucosyltransferase genes among strains of
Streptococcus mutans";
RL FEMS Microbiol. Lett. 161:331-336(1998).
DR EMBL; D88658; BAA26110.1; -;
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 10.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Transference.
SQ SEQUENCE 1455 AA; 163046 MW; 6D90A4978D35DD82 CRC64;

Query Match 0.8%; Score 15; DB 2; Length 1455;
Best Local Similarity 100.0%; Pred. No. 9.5e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 986 DNSNPVQAEQLNWL 1000
Db 439 DNSNPVQAEQLNWL 453

RESULT 10
O69391 PRELIMINARY; PRT; 1455 AA.
AC O69391;
DT 01-AUG-1998 (TRENBLrel. 07, Created)
DT 01-AUG-1998 (TRENBLrel. 07, Last sequence update)
DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
DE GLUCOSYLTRANSFERASE-SI.
GN GTFC.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
ON Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-WT4251;
RX MEDLINE-98231643; PubMed-9570124;
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,
Kimura S., Hamada S.;
RT "Molecular analyses of glucosyltransferase genes among strains of
Streptococcus mutans";
RL FEMS Microbiol. Lett. 161:331-336(1998).
DR EMBL; D88661; BAA26114.1; -;
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 9.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Transference.
SQ SEQUENCE 1455 AA; 162804 MW; 683A359D873E9E1A CRC64;

Query Match 0.8%; Score 15; DB 2; Length 1455;
Best Local Similarity 100.0%; Pred. No. 9.5e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 9
O69388 PRELIMINARY; PRT; 1455 AA.
AC O69388;
DT 01-AUG-1998 (TRENBLrel. 07, Created)
DT 01-AUG-1998 (TRENBLrel. 07, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE GLUCOSYLTRANSFERASE-SI.
GN GTFC.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
ON Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-WT4239;
RX MEDLINE-98231643; PubMed-9570124;
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,
Kimura S., Hamada S.;
RT "Molecular analyses of glucosyltransferase genes among strains of
Streptococcus mutans";
RL FEMS Microbiol. Lett. 161:331-336(1998).
DR EMBL; D88661; BAA26114.1; -;
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 9.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Transference.
SQ SEQUENCE 1455 AA; 162804 MW; 683A359D873E9E1A CRC64;

QY 986 DNSNPVQAEQLNWL 1000
|||||
Db 439 DNSNPVQAEQLNWL 453

RESULT 11

O69397 ID O69397 PRELIMINARY; PRT; 1455 AA.
AC O69397
DT 01-AUG-1998 (TREMELrel. 07, Created)
DT 01-AUG-1998 (TREMELrel. 07, Last sequence update)
DT 01-DEC-2001 (TREMELrel. 19, Last annotation update)
DE GLUCOSYLTRANSFERASE-SI.
GN GTEC.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-WT4467;
RX MEDLINE-98231643; PubMed-9570124;
RA Fujiwara T., Terada Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,
Kimura S., Hamada S.;
RT "Molecular analyses of glucosyltransferase genes among strains of
Streptococcus mutans.";
RL FEMS Microbiol. Lett. 161:331-336(1998).
DR EMBL; D89978; BAA26120.1; -
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 9.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Transferase.
SQ SEQUENCE 1455 AA; 162914 MW; A1263427BF24E8E1 CRC64;

Query Match 0.8%; Score 15; DB 2; Length 1455;
Best Local Similarity 100.0%; Pred. No. 9.5e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 986 DNSNPVQAEQLNWL 1000
|||||
Db 439 DNSNPVQAEQLNWL 453

RESULT 12

O9L466 ID O9L466 PRELIMINARY; PRT; 1477 AA.
AC O9L466
DT 01-OCT-2000 (TREMELrel. 15, Created)
DT 01-OCT-2000 (TREMELrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMELrel. 19, Last annotation update)
DE DEXTRANSUCRASE (EC 2.4.1.5).
GN DSRG.
OS Leuconostoc mesenteroides.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
OC Leuconostoc.
OX NCBI_TaxID=1245;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NRRL B-1355;
RA Arguello-Morales M.A., Remaud-Simeon M., Pizzut S., Sarcabal P.,
Willemot R.M., Monsan P.;
RT "Sequence analysis of the gene encoding alternansucrase, a sucrose
glucosyltransferase from Leuconostoc mesenteroides NRRL B-1355.";
RL Submitted (OCT-1999) to the EMBL/Genbank/DDBJ databases.
DR EMBL; AJ250172; CAB75565.1; -
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 14.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Transferase; Glycosyltransferase.
SQ SEQUENCE 1477 AA; 164887 MW; B6F5710DEDFCB831 CRC64;

Query Match 0.8%; Score 15; DB 2; Length 1477;
Best Local Similarity 100.0%; Pred. No. 9.6e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 986 DNSNPVQAEQLNWL 1000
|||||
Db 464 DNSNPVQAEQLNWL 478

RESULT 13

Q9EZHS ID Q9EZHS PRELIMINARY; PRT; 1508 AA.
AC Q9EZHS;
DT 01-MAR-2001 (TREMELrel. 16, Created)
DT 01-MAR-2001 (TREMELrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMELrel. 17, Last annotation update)
DE DEXTRANSUCRASE DSRB742.
GN DSRB742.
OS Leuconostoc mesenteroides.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
OC Leuconostoc.
OX NCBI_TaxID=1245;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-B-742CB;
RA Kim H.-S., Kim D., Ryu H.-J., Robyt J.F.;
RT "Leuconostoc mesenteroides B-742CB, a dextranucrase gene.";
RL Submitted (AUG-2000) to the EMBL/Genbank/DDBJ databases.
DR EMBL; AF294469; AAG38021.1; -
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 14.
DR Pfam; PF02324; Glyco_hydro_70; 1.
SQ SEQUENCE 1508 AA; 168542 MW; E2FCFA0F87AE4F3A CRC64;

Query Match 0.8%; Score 15; DB 2; Length 1508;
Best Local Similarity 100.0%; Pred. No. 9.8e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 986 DNSNPVQAEQLNWL 1000
|||||
Db 495 DNSNPVQAEQLNWL 509

RESULT 14

Q59983 ID Q59983 PRELIMINARY; PRT; 1590 AA.
AC Q59983;
DT 01-NOV-1996 (TREMELrel. 01, Created)
DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMELrel. 19, Last annotation update)
DE GLUCOSYLTRANSFERASE-I PRECURSOR (EC 2.4.1.5).
GN GTFI.
OS Streptococcus sobrinus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1310;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-OM2176;
RX MEDLINE-94146405; PubMed-8312602;
RA Sato S., Inoue M., Handa N., Aizawa Y., Isobe Y., Katayama T.;
RT "DNA sequence of the glucosyltransferase gene of serotype d
Streptococcus sobrinus.";
RL DNA Seq. 4:19-27(1993).
DR EMBL; D13858; BAA02376.1; -
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 16.
DR Pfam; PF02324; Glyco_hydro_70; 1.

KW Signal; Transferase; Glycosyltransferase.
 FT SIGNAL 1 38 POTENTIAL.
 FT CHAIN 39 1590 GLUCOSYLTRANSFERASE-I.
 SQ SEQUENCE 1590 AA; 175955 MW; C3C83A57CF3C2B0E CRC64;

Query Match 0.8%; Score 15; DB 2; Length 1590;
 Best Local Similarity 100.0%; Pred. No. 1e-05;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 986 DNSNPVYQAEQLNWL 1000
 Db 409 DNSNPVYQAEQLNWL 423

RESULT 15
 Q55263
 ID Q55263 PRELIMINARY; PRT; 1590 AA.
 AC Q55263;
 DT 01-NOV-1996 (TRENBLrel. 01, Created)
 DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)
 DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
 DE GTF-I.
 GN GLUCOSYLTRANSFERASE.
 OS Streptococcus sobrinus.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1310;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 33478;
 RA Sato S.;
 RT "DNA and amino-acid sequences of water-insoluble-glucan synthetase
 RT produced from Streptococcus sobrinus ATCC 33478.";
 RL Ann. Kagoshima Univ. Dental School 16:23-29(1996).
 DR EMBL; D63570; BAA09792.1;
 DR InterPro; IPR002479; CW_binding.
 DR InterPro; IPR003318; Glyco_hydro_70.
 DR Pfam; PF01473; CW_binding_1; 15.
 DR Pfam; PF02324; Glyco_hydro_70; 1.
 SQ SEQUENCE 1590 AA; 176058 MW; 9DF7A3F2C6B4FD43 CRC64;

Query Match 0.8%; Score 15; DB 2; Length 1590;
 Best Local Similarity 100.0%; Pred. No. 1e-05;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 986 DNSNPVYQAEQLNWL 1000
 Db 409 DNSNPVYQAEQLNWL 423

RESULT 16
 O68542
 ID O68542 PRELIMINARY; PRT; 1449 AA.
 AC O68542;
 DT 01-AUG-1998 (TRENBLrel. 07, Created)
 DT 01-AUG-1998 (TRENBLrel. 07, Last sequence update)
 DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
 DE GLUCOSYLTRANSFERASE N (FRAGMENT).
 GN GTFN.
 OS Streptococcus salivarius.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1304;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-V1477;
 RA Jaffe R.I.;
 RT "Streptococcus salivarius V1477 gtfN.";
 RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF049609; AAC05156.1;
 DR InterPro; IPR002479; CW_binding.

DR InterPro; IPR003318; Glyco_hydro_70.
 DR Pfam; PF01473; CW_binding_1; 8.
 DR Pfam; PF02324; Glyco_hydro_70; 1.
 KW Transferase.
 FT NON_TER 1449 1449
 SQ SEQUENCE 1449 AA; 159895 MW; 0700F6D748471BFB CRC64;

Query Match 0.7%; Score 13; DB 2; Length 1449;
 Best Local Similarity 100.0%; Pred. No. 0.0011;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 978 EFLANDIDNSNP 990
 Db 460 EFLANDIDNSNP 472

RESULT 17
 Q55264
 ID Q55264 PRELIMINARY; PRT; 1449 AA.
 AC Q55264;
 DT 01-NOV-1996 (TRENBLrel. 01, Created)
 DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
 DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
 DE GLUCOSYLTRANSFERASE PRECURSOR.
 GN GTFL.
 OS Streptococcus salivarius.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1304;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-95122197; PubMed-7822030;
 RA Simpson C.L., Giffard P.M., Jacques N.A.;
 RT "Streptococcus salivarius ATCC 25975 possesses at least two genes
 RT coding for primer-independent glucosyltransferases.";
 RL Infect. Immun. 63:609-621(1995).
 DR EMBL; L35495; AAC41412.1;
 DR InterPro; IPR002479; CW_binding.
 DR InterPro; IPR003318; Glyco_hydro_70.
 DR Pfam; PF01473; CW_binding_1; 8.
 DR Pfam; PF02324; Glyco_hydro_70; 1.
 KW Signal; Transferase.
 FT SIGNAL 1 35 POTENTIAL.
 FT CHAIN 36 1449 GLUCOSYLTRANSFERASE.
 SQ SEQUENCE 1449 AA; 159984 MW; DD62F07306E86A46 CRC64;

Query Match 0.7%; Score 13; DB 2; Length 1449;
 Best Local Similarity 100.0%; Pred. No. 0.0011;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 978 EFLANDIDNSNP 990
 Db 460 EFLANDIDNSNP 472

RESULT 18
 O52224
 ID O52224 PRELIMINARY; PRT; 1508 AA.
 AC O52224;
 DT 01-JUN-1998 (TRENBLrel. 06, Created)
 DT 01-JUN-1998 (TRENBLrel. 06, Last sequence update)
 DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
 DE GLUCOSYLTRANSFERASE (EC 2.4.1.5).
 GN DSRB.
 OS Leuconostoc mesenteroides.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
 OC Leuconostoc.
 OX NCBI_TaxID=1245;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-NREL B-1299;

RA Monchois V., Renaud-Simeon M., Monsan P., Willemot R.M.;
 RT "Cloning and sequencing of a gene coding for an extracellular
 RT dextranucrase (DSR) from Leuconostoc mesenteroides NRRL B-1299
 RT synthesizing only a (1-6) glucan.";
 RL FEMS Microbiol. Lett. 0:0-0(1998).
 DR EMBL; AF030129; AAB95453.1; -;
 DR InterPro; IPR002479; CW_binding.
 DR Pfam; PF01473; CW_binding_1; 14.
 DR Pfam; PF02324; Glyco_hydro_70; 1.
 KW Transferase; Glycosyltransferase.
 SQ SEQUENCE 1508 AA; 168511 MW; E70CECB57A70D1F0 CRC64;

Query Match 0.7%; Score 13; DB 2; Length 1508;
 Best Local Similarity 100.0%; Pred. No. 0.0011;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1508 ADMVPDQIYNLPG 1520
 DB 1004 ADMVPDQIYNLPG 1016

RESULT 19

ID Q00600 PRELIMINARY; PRT; 1518 AA.
 AC Q00600;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE GLUCOSYLTRANSFERASE I (EC 2.4.1.5) (GTF) (DEXTRANSUCRASE).
 DE (SUCROSE 6-GLUCOSYLTRANSFERASE).
 GN GTFI.
 OS Streptococcus salivarius.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1304;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 25975;
 RX MEDLINE=92148377; PubMed=1838391;
 RA Giffard P.M., Simpson C.L., Milward C.P., Jacques N.A.;
 RT "Molecular characterization of a cluster of at least two
 RT glucosyltransferase genes in Streptococcus salivarius ATCC 25975.";
 RL J. Gen. Microbiol. 137:2577-2593(1991).
 CC -!- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS. THAT ARE THOUGHT TO
 CC PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE OF
 CC THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE
 CC AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.
 CC -!- CATALYTIC ACTIVITY: SUCROSE + ((1,6)-ALPHA-D-GLUCOSYL)(N) - D-
 CC FRUCTOSE + ((1,6)-ALPHA-D-GLUCOSYL)(N+1).
 CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
 CC -!- DISEASE: DENTAL CARIES.
 DR EMBL; Z11873; CAA77900.1; -;
 DR EMBL; M64111; AAA26896.1; -;
 DR InterPro; IPR002479; CW_binding.
 DR InterPro; IPR003318; Glyco_hydro_70.
 DR Pfam; PF01473; CW_binding_1; 13.
 DR Pfam; PF02324; Glyco_hydro_70; 1.
 KW Transferase; Glycosyltransferase; Repeat; Dental caries.
 FT DOMAIN 1307 1482 6 DIRECT REPEATS.
 FT REPEAT 1307 1338 REPEAT 1.
 FT REPEAT 1339 1352 REPEAT 2.
 FT REPEAT 1372 1403 REPEAT 3.
 FT REPEAT 1404 1417 REPEAT 4.
 FT REPEAT 1437 1468 REPEAT 5.
 FT REPEAT 1469 1482 REPEAT 6.
 SQ SEQUENCE 1518 AA; 167730 MW; DAA41E717098B59A CRC64;

Query Match 0.7%; Score 13; DB 2; Length 1518;
 Best Local Similarity 100.0%; Pred. No. 0.0012;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1371 AWWVPVGASDNQD 1383
 DB 858 AWWVPVGASDNQD 870

RESULT 20

ID Q92AR4 PRELIMINARY; PRT; 1527 AA.
 AC Q92AR4;
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE DEXTRANSUCRASE.
 GN DEX.
 OS Leuconostoc mesenteroides.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
 OC Leuconostoc.
 OX NCBI_TaxID=1245;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NRRL B-512-F;
 RA Bhatnagar R., Singh D.K.S.;
 RT "Cloning and Molecular Characterization of Dextranucrase Gene from
 RT Leuconostoc mesenteroides NRRL B-512F.";
 RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U81374; RAD10952.1; -;
 DR InterPro; IPR002479; CW_binding.
 DR InterPro; IPR003318; Glyco_hydro_70.
 DR Pfam; PF01473; CW_binding_1; 16.
 DR Pfam; PF02324; Glyco_hydro_70; 1.
 SQ SEQUENCE 1527 AA; 169709 MW; 1DFAFA237C743398 CRC64;

Query Match 0.7%; Score 13; DB 2; Length 1527;
 Best Local Similarity 100.0%; Pred. No. 0.0012;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1366 VSGILAVWVPVGA 1378
 DB 881 VSGILAVWVPVGA 893

RESULT 21

ID Q9RE05 PRELIMINARY; PRT; 2057 AA.
 AC Q9RE05;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE ALTERNANSUCRASE (EC 2.4.1.140).
 GN ASR.
 OS Leuconostoc mesenteroides.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
 OC Leuconostoc.
 OX NCBI_TaxID=1245;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NRRL B-1355;
 RX MEDLINE=20080809; PubMed=10612736;
 RA Arguello-Morales M.A., Renaud-Simeon M., Pizzut S., Sarcabal P.,
 RA Willemot R.M., Monsan P.;
 RT "Sequence analysis of the gene encoding alternansucrase, a sucrose
 RT glucosyltransferase from Leuconostoc mesenteroides NRRL B-1355.";
 RL FEMS Microbiol. Lett. 182:81-85(2000).
 DR EMBL; AJ250173; CAB65910.2; -;
 DR InterPro; IPR006004; Chlamydia_OMP.
 DR InterPro; IPR002479; CW_binding.
 DR InterPro; IPR003318; Glyco_hydro_70.
 DR Pfam; PF01473; CW_binding_1; 12.
 DR Pfam; PF02324; Glyco_hydro_70; 1.
 DR ProDom; PD001717; Chlamydia_OMP; 1.
 KW Transferase; Glycosyltransferase.

SQ SEQUENCE 2057 AA; 228987 MW; 62BCE9385D9A11BE CRC64;

Query Match 0.7%; Score 13; DB 2; Length 2057;
Best Local Similarity 100.0%; Pred. No. 0.0015;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 978 EFLANDIDNSNP 990
Db 589 EFLANDIDNSNP 601
|||||

RESULT 22
Q9WXJ4 PRELIMINARY; PRT; 1338 AA.

ID Q9WXJ4
AC Q9WXJ4;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE GTF-S.
CN GTF-S.
OS Streptococcus criceti.
OC Plasmid pAM1.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OC NCBI_TaxID=1333;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HS-6;
RA Inoue M., Fukui K., Miyagi A.;
RT "S.cricetus glucosyltransferase(gtfs and gtf) genes.";
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB026123; BAA77236.1; -;
DR InterPro; IPR002479; CW_Binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 10.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Plasmid.
SQ SEQUENCE 1338 AA; 148558 MW; 0A90C8E10F15D99B CRC64;

Query Match 0.6%; Score 11; DB 2; Length 1338;
Best Local Similarity 100.0%; Pred. No. 0.12;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1251 ILTSVRKXGKA 1261
Db 640 ILTSVRKXGKA 650
|||||

RESULT 23
Q9WXJ5 PRELIMINARY; PRT; 1512 AA.

ID Q9WXJ5
AC Q9WXJ5;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE GTF-S.
CN GTF-S.
OS Streptococcus criceti.
OC Plasmid pAM1.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OC NCBI_TaxID=1333;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HS-6;
RA Inoue M., Fukui K., Miyagi A.;
RT "S.cricetus glucosyltransferase(gtfs and gtf) genes.";
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB026123; BAA77237.1; -;
DR InterPro; IPR002479; CW_Binding.
DR InterPro; IPR003318; Glyco_hydro_70.

DR Pfam; PF01473; CW_binding_1; 14.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Plasmid.
SQ SEQUENCE 1512 AA; 167145 MW; 4C03D9C8C601FC14 CRC64;

Query Match 0.6%; Score 10; DB 2; Length 1512;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 986 DNSNPVQAE 995
Db 417 DNSNPVQAE 426
|||||

RESULT 24
Q9R4L7 PRELIMINARY; PRT; 20 AA.

ID Q9R4L7
AC Q9R4L7;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE DEXTRANSUCRASE (EC 2.4.1.5) (FRAGMENT).
OS Leuconostoc mesenteroides.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
OC Leuconostoc.
OC NCBI_TaxID=1245;
RN [1]
RP SEQUENCE.
RX MEDLINE=95306915; PubMed=7540436;
RA Funane K., Yamada M., Shiraiwa N., Takahara H., Yamamoto N.,
RA Ichishima E., Kobayashi M.;
RT "Aggregated form of dextranucrases from Leuconostoc mesenteroides
NRRL B-512F and its constitutive mutant.";
RL Biosci. Biotechnol. Biochem. 59:776-780(1995).
SQ SEQUENCE 20 AA; 2356 MW; 701179A7AC89557F CRC64;

Query Match 0.5%; Score 9; DB 2; Length 20;
Best Local Similarity 100.0%; Pred. No. 0.39;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1198 PRVYGDLY 1206
Db 4 PRVYGDLY 12
|||||

RESULT 25
O18799 PRELIMINARY; PRT; 100 AA.

ID O18799
AC O18799;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE GLYCOPHORIN A PRECURSOR (FRAGMENT).
GN GPA.
OS Gorilla gorilla (gorilla).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Gorilla.
OC NCBI_TaxID=9593;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IB;
RX MEDLINE=97381163; PubMed=9238519;
RA Xie S.-S., Huang C.-H., Reid M.E., Blancher A., Blumenfeld O.O.;
RT "The glycophorin A gene family in gorillas: structure, expression, and
RT comparison with the human and chimpanzee homologues.";
RL Biochem. Genet. 35:59-76(1997).
DR EMBL: AF015175; AAB81212.1; -;
DR HSSP; P02724; IAF0.
DR InterPro; IPR001195; Glycophorin_A.
DR Pfam; PF01102; Glycophorin_A; 1.
DR PROSITE; PS00312; GLYCOPHORIN_A; 1.

KW Signal.
FT NON_TER 1 1
FT SIGNAL <1 12 POTENTIAL.
FT CHAIN 13 >100 GLYCOPHORIN A.
FT NON_TER 100 100
SQ SEQUENCE 100 AA; 10720 MW; 22E0BCC0095AC77 CRC64;

Query Match 0.4%; Score 8; DB 6; Length 100;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 695 VOLVHRFS 702
Db 42 VOLVHRFS 49

RESULT 26
Q18797
ID Q18797 PRELIMINARY; PRT; 123 AA.
AC Q18797;
DT 01-JAN-1998 (TREMELrel. 05, Created)
DT 01-JAN-1998 (TREMELrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMELrel. 19, Last annotation update)
DE GLYCOPHORIN A PRECURSOR (FRAGMENT).
GN GPA.
OS Gorilla gorilla (Gorilla).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Gorilla.
OX NCBI_TaxID=9593;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1A;
RX MEDLINE=97381163; PubMed=9238519;
RA Xie S.-S., Huang C.-H., Reid M.E., Blancher A., Blumenfeld O.O.;
RT "The glycophorin A gene family in gorillas: structure, expression, and
comparison with the human and chimpanzee homologues.";
RL Blochem. Genet. 35:59-76(1997).
DR EMBL; AF015173; AAB81210.1; .
DR HSSP; P02724; 1AFO.
DR InterPro; IPR001195; Glycophorin_A.
DR Pfam; PF01102; Glycophorin_A; 1.
DR PROSITE; PS00312; GLYCOPHORIN_A; 1.
KW Signal.
FT NON_TER 1 1
FT SIGNAL <1 11 POTENTIAL.
FT CHAIN 12 >123 GLYCOPHORIN A.
FT NON_TER 123 123
SQ SEQUENCE 123 AA; 13609 MW; 4A595A6C25C1CD19 CRC64;

Query Match 0.4%; Score 8; DB 6; Length 123;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 695 VOLVHRFS 702
Db 73 VOLVHRFS 80

RESULT 27
Q9CNV9
ID Q9CNV9 PRELIMINARY; PRT; 196 AA.
AC Q9CNV9;
DT 01-JUN-2001 (TREMELrel. 17, Created)
DT 01-JUN-2001 (TREMELrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMELrel. 19, Last annotation update)
DE HYPOTHETICAL PROTEIN PM0314.
GN PM0314.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Pasteurella.
OX NCBI_TaxID=747;

[1]
RN SEQUENCE FROM N.A.
RP STRAIN=PM70;
RX MEDLINE=21145866; PubMed=11248100;
RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RT "Complete genomic sequence of Pasteurella multocida Pm70.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
DR EMBL; AEO06067; AAK02398.1; .
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 196 AA; 22954 MW; 39204C50D34C6377 CRC64;

Query Match 0.4%; Score 8; DB 16; Length 196;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 784 QFAKGTVS 791
Db 147 QFAKGTVS 154

RESULT 28
Q9KYA1
ID Q9KYA1 PRELIMINARY; PRT; 213 AA.
AC Q9KYA1;
DT 01-OCT-2000 (TREMELrel. 15, Created)
DT 01-OCT-2000 (TREMELrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMELrel. 17, Last annotation update)
DE PUTATIVE INTEGRAL MEMBRANE PROTEIN.
GN SCK15.03.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Seeger K.J., Harris D.;
RL Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapalte D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL; AL356813; CAB92594.1; .
DR InterPro; IPR003439; ABC transporter.
DR PROSITE; PS00211; ABC_TRANSPORTER; UNKNOWN_1.
SQ SEQUENCE 213 AA; 23483 MW; 7800326A3752DE66 CRC64;

Query Match 0.4%; Score 8; DB 2; Length 213;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 254 NGELHATG 261
Db 164 NGELHATG 171

RESULT 29
O53780
ID O53780 PRELIMINARY; PRT; 228 AA.
AC O53780;
DT 01-JUN-1998 (TREMELrel. 06, Created)

DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE P[UTATIVE LIPOPROTEIN.
GN LPON OR RV0583C OR MT039.21C.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37KV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
DR EMBL; AL021942; CAA17454.1; -.
DR Tuberculist; RV0583c; -.
KW Lipoprotein; Complete proteome.
SQ SEQUENCE 228 AA; 23682 MW; 1F0821C9E9947526 CRC64;

Query Match 0.4%; Score 8; DB 16; Length 228;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 TAAVATVA 25
Db 5 TAAVATVA 12
|||||||

RESULT 30
O65770 PRELIMINARY; PRT; 292 AA.
ID O65770
AC O65770;
DT 01-AUG-1998 (Tremblrel. 07, Created)
DT 01-AUG-1998 (Tremblrel. 07, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE HOMEODOMAIN LEUCINE ZIPPER PROTEIN.
GN HB-2.
OS Craterostigma plantagineum.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Lamiales; Lamiales incertae sedis; Torenieae;
OC Craterostigma.
OX NCBI_TaxID=4153;
RN [1]
RP SEQUENCE FROM N.A.
RA Frank W., Phillips J., Salamini F., Bartels D.;
RT "Two dehydration-inducible transcripts from the resurrection plant
RT Craterostigma plantagineum encode interacting homeodomain-leucine
RT zipper proteins.";
RL Plant J. 15:413-421(1999).
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -!- SIMILARITY: WITH OTHER HOMEODOMAIN PROTEINS.
DR EMBL; AJ005833; CNA06728.1; -.
DR HSP; P01366; 1AKH.
DR TRANSFAC; T04168; -.
DR InterPro; IPR003106; HALZ.
DR InterPro; IPR001356; Homeobox.
DR Pfam; PF02183; HALZ; 1.
DR Pfam; PF00046; homeobox; 1.
DR SMART; SM00340; HALZ; 1.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEODOMAIN_1; 1.
DR PROSITE; PS00071; HOMEODOMAIN_2; 1.

KW DNA-binding; Homeobox; Nuclear protein.
SQ SEQUENCE 292 AA; 32109 MW; C607BA8260A25BFB CRC64;

Query Match 0.4%; Score 8; DB 10; Length 292;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 309 VGAANS GF 316
Db 266 VGAANS GF 273
|||||||

Search completed: August 12, 2002, 08:16:21
Job time: 362 sec